ESTIMATION OF STATE VARIABLES FOR NOISY DYNAMIC SYSTEMS

BY

HENRY COX

MAY 1963

Doctor of Science Thesis
Electrical Engineering Department
Massachusetts Institute of Technology

ESTIMATION OF STATE VARIABLES

FOR NOISY DYNAMIC SYSTEMS

by

HENRY COX

Lieutenant, United States Navy

S.B., College of the Holy Cross

(1956)

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1963

ESTIMATION OF STATE VARIABLES FOR NOISY DYNAMIC SYSTEMS

by

HENRY COX

Lieutenant, United States Navy

Submitted to the Department of Electrical Engineering on May 10, 1963, in partial fulfillment of the requirements for the degree of Doctor of Science.

ABSTRACT

The problem of estimating state variables and parameters of dynamic systems in the presence of random disturbances and measurement noise is examined for a wide range of dynamic systems, including linear and nonlinear systems in discrete- and continuous-time. This is a fundamental statistical problem arising in, but by no means limited to, the areas of adaptive and optimum control. Parameter estimation is treated as a special case of state variable estimation.

The general solution of the linear problem is given. Approximation techniques are developed for the nonlinear problem and the results of simulation studies demonstrating the application of these techniques are presented.

A dynamic programming formulation of the estimation problem is developed.

A discussion of the concepts of controllability and observability is given in which particular attention is paid to the problems peculiar to discrete-time systems.

Background material on matrix identities, Gaussian random vectors, the generalized inverse of a matrix, and probability density functionals is included in a series of appendices.

Thesis Supervisor: Ronald A. Howard

Title: Associate Professor of Electrical Engineering

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Professor Ronald A. Howard for his continuing guidance, encouragement and assistance throughout the course of this research. The author is very grateful to Professor Lofti A. Zadeh of the University of California at Berkeley for many stimulating conversations and helpful suggestions during his stay at M.I.T. in the fall of 1962. Thanks are due to Professors L.A. Gould and C.E. Shannon who participated in the supervision of this thesis.

The author is indebted to the Bureau of Ships, United States Navy, for his financial support and to its representatives, Captain Edward S. Arentzen, U.S.N. (Ret.), Captain William M. Nicholson, U.S.N. and Commander John R. Baylis, U.S.N. for encouragement through the course of this program.

The hundreds of stimulating discussions with his colleague, Lieutenant R.B. Gilchrist, U.S.N., have been a significant part of the author's stay at M.I.T.

The dynamic programming formulation of the discrete-time estimation problem is an outgrowth of a discussion with J.D.R. Kramer, Jr. The simulation studies reported in Chapter V were carried out at the Computation Center, M.I.T.

Finally, the author wishes to thank his wife, Mary Ann, for her expert typing, invaluable assistance in editing and, most of all, for her willing acceptance of the many sacrifices imposed by the doctoral program.

TABLE OF CONTENTS

TABLE OF CONTENTS (CON'T)

TABLE OF CONTENTS (CON'T)

INTRODUCTION

The introduction of statistical considerations into a wide range of engineering problems is a logical consequence of the existence of uncertainties of one type or another which frequently make a purely deterministic approach to these problems unrealistic. The field of automatic control abounds in these uncertainties, such as those arising from random disturbances, measurement noise and changing environment. Often uncertainties about the process or plant being controlled and about future commands are also present. Hence, it is not surprising that statistical methods have contributed significantly to the development of modern control theory.

A fundamental problem in the areas of adaptive and optimum control is that of estimation of state variables and parameters of a dynamic system when random disturbances and measurement noise are present. If the estimates of these quantities are to be used for control purposes, then the estimation problem must be solved in real time. The problem then is to synthesize an estimator which will act on all available data and continually produce up-to-date estimates of the quantities of interest. In related problems of <u>ex post facto</u> data analysis the severe real time requirement is relaxed and, at the expense of computation time, priority is given to processing the often limited amount of data available in an optimal manner.

This report is devoted to the examination of this estimation problem for a wide range of dynamic systems. Chapter I presents the mathematical model to be used in the continuous-time version of the problem and the generality of the model and its relation to other problems is discussed. Because the analogous discrete-time problem is conceptually

simpler, we first examine it in detail in Chapter II before returning to the continuous-time problem in Chapter III. Chapter IV discusses the concepts of controllability and observability. Chapter V gives the results of some computational studies; while several concluding remarks and suggestions for possible further research are contained in Chapter VI. Material of a complementary and supplementary nature is included in a series of appendices.

The principal references for this report are the work of Kalman [29, 31, 32, 34] and the paper by Bryson and Frazier [10]. An effort has been made to use a notation which is compatible with these references.

CHAPTER I

BACKGROUND AND PERSPECTIVE

1.1 Introduction

The purpose of this chapter is to lay the groundwork for the analysis of the estimation problem and to discuss various aspects of the mathematical models considered in the sequel. After taking care of some notational preliminaries and reviewing briefly the state approach to dynamic systems, the continuous-time model is presented. This model is then related to control problems and to classical linear filtering theory. The author chose the continuous-time model for this preliminary discussion because it is easily related to the model appearing most frequently in modern control theory.

1.2 Notation

For simplicity vector notation is used throughout this work. All vectors are column vectors. Vectors are designated by underlined lower case letters and matrices by underlined upper case letters. For example, the vector $\underline{x}$ and the matrix $\underline{A}$ are to be interpreted as follows:

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \qquad \underline{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix}$$

The transpose of $\underline{x}$ and $\underline{A}$ are denoted by $\underline{x}'$ and $\underline{A}'$ respectively. The inner product of two vectors $\underline{x}$ and $\underline{y}$ is written as a matrix product. That is,

$$\underline{x}'\underline{y} = \underline{y}'\underline{x} = \sum_i x_i y_i$$

The Euclidian norm of a vector $\underline{x}$ is denoted by $\|\underline{x}\|$.

$$\|\underline{x}\| = \sqrt{\underline{x}'\underline{x}} = \sqrt{x_1^2 + \cdots + x_n^2}$$

Also

$$\|\underline{x}\|^2 = \underline{x}'\underline{x} = \sum_i x_i^2$$

If $\underline{A}$ is a positive definite matrix, the following special notation is used for the associated quadratic form

$$\|\underline{x}\|_{\underline{A}}^2 = \underline{x}'\underline{A}\underline{x} = \sum_{ij} a_{ij} x_i x_j$$

If $\underline{x}$ is a function of time, then $\dot{\underline{x}}(t)$ is the derivative of $\underline{x}$ with respect to time. That is,

$$\dot{\underline{x}}(t) = d\underline{x}(t)/dt = \begin{bmatrix} dx_1(t)/dt \\ dx_2(t)/dt \\ \cdot \\ \cdot \\ \cdot \\ dx_n(t)/dt \end{bmatrix}$$

The scalar function $V(x_1(t),\ldots,x_n(t);t)$ is written $V(\underline{x}(t),t)$. The gradient of V is a vector denoted by $V_{\underline{x}}$. The partial derivative of

V with respect to t is written V_t. That is,

$$V_{\underline{x}} = \text{grad } V = \begin{bmatrix} \partial V/\partial x_1 \\ \cdot \\ \cdot \\ \cdot \\ \partial V/\partial x_n \end{bmatrix} \qquad V_t = \partial V(\underline{x}(t),t)/\partial t$$

Vector valued functions are designated by underlined lower case letters. For example,

$$\underline{f}(\underline{x},\underline{u},t) = \begin{bmatrix} f_1(x_1,\ldots,x_n,u_1,\ldots,u_r,t) \\ \cdot \\ \cdot \\ \cdot \\ f_n(x_1,\ldots,x_n,u_1,\ldots,u_r,t) \end{bmatrix}$$

We use $(\partial \underline{f}(\underline{x}_o)/\partial \underline{x})$ to designate the Jacobian matrix of partial derivatives $\{\partial f_i/\partial x_j\}$ evaluated at $\underline{x} = \underline{x}_o$.

The segment of the time-function $\underline{u}(t)$ on the closed interval $t_o \leq t \leq t_1$ is designated by $\underline{u}_{[t_o,t_1]}$.

1.3 Equations of State

The convention of describing a dynamic system by a set of first order ordinary differential (or difference) equations or, more compactly, by a vector differential (or difference) equation is already well established in the control literature. Introductory accounts of "state-space techniques", the name usually given to a body of concepts associated with this characterization, are given in references [7,19,68]. However, since vector notation will be used throughout this report, it is appropriate to review briefly some of the basic ideas and relations.

Informally, we may say that the future behavior of a system depends upon its present state and any forcing functions or inputs which may be applied to the system in the future. We limit our discussion to systems in which the minimum number of quantities necessary to specify the state of the system at time t is finite. The members $x_1(t),\ldots,x_n(t)$ of such a minimum finite set are called state variables. The state variables are elements of the state vector $\underline{x}(t)$ which ranges over a set X, called the state space. Examples of appropriate sets of state variables are the amount of charge on each capacitor and the amount of current in each inductor in an electric circuit, the position and momentum of each mass of a mechanical system and the output of each integrator of an analogue computer representation of a system.

The equations of state for a continuous-time system are usually written as a system of first order ordinary differential equations of the form

$$\mathring{x}_i(t) = f_i(x_1(t),\ldots,x_n(t),u_1(t),\ldots,u_r(t),t) \qquad i = 1,\ldots,n$$

where $u_1,\ldots,u_r$ are the inputs of the system. The equivalent vector equation is

$$\mathring{\underline{x}}(t) = \underline{f}(\underline{x}(t),\underline{u}(t),t) \tag{1.1}$$

The analogous equations for a discrete-time system are

$$x_i(k+1) = f_i(x_1(k),\ldots,x_n(k),u_1(k),\ldots,u_r(k),k) \qquad i = 1,\ldots,n$$

and

$$\underline{x}(k+1) = \underline{f}(\underline{x}(k),\underline{u}(k),k) \tag{1.2}$$

While equations (1.1) and (1.2) encompass a very large class of systems, it is well to remember that they do not include distributed parameter systems described by partial differential equations, systems with time delays described by differential-difference equations and other systems of more complicated types.

As an example of the reduction of an nth order nonlinear differential equation to a system of n first order differential equations, consider the equation

$$y^{(n)}(t) + g(y,\dot{y},\ldots,y^{(n-1)},u,t) = 0$$

Let $x_1 = y$, $x_2 = \dot{y}$, . . . , $x_n = y^{(n-1)}$. Then an equivalent system of first order equations is

$$\dot{x}_1(t) = x_2(t)$$
$$\dot{x}_2(t) = x_3(t)$$
$$\vdots$$
$$\dot{x}_{n-1}(t) = x_n(t)$$
$$\dot{x}_n(t) = -g(x_1,\ldots,x_n,u,t)$$

Important special cases of (1.1) and (1.2) arise when the system is linear. For a linear system (1.1) becomes

$$\underline{\dot{x}}(t) = \underline{F}(t)\underline{x}(t) + \underline{G}(t)\underline{u}(t) \tag{1.3}$$

One can easily verify that the solution of (1.3) has the following form;

$$\underline{x}(t) = \underline{\Phi}(t,t_o)\underline{x}(t_o) + \int_{t_o}^{t} \underline{\Phi}(t,s)\underline{G}(s)\underline{u}(s)\underline{ds}$$

where $\underline{\Phi}(t,t_o)$ is the transition or fundamental matrix of the system (1.3). The transition matrix is the unique matrix satisfying the following relations;

$$\frac{\partial}{\partial t} \underline{\Phi}(t,t_o) = \underline{F}(t)\,\underline{\Phi}(t,t_o)$$

$$\underline{\Phi}(t_o,t_o) = \underline{I}$$

where $\underline{I}$ is the identity matrix. For the special case of time-invariant linear systems, $\underline{F}$ is a constant matrix and the transition matrix is given by

$$\underline{\Phi}(t,t_o) = \exp\left[\,(t-t_o)\underline{F}\,\right] = \sum_{n=o}^{\infty} \frac{(t-t_o)^n \underline{F}^n}{n!}$$

As an example of the reduction of a linear differential equation to a system of first order equations of the form of (1.3), consider the third order time-invariant equation

$$\dddot{y}(t) + a_2\,\ddot{y}(t) + a_1\,\dot{y}(t) + a_o\,y(t) = b_2\ddot{u}(t) + b_1\dot{u}(t) + b_o u(t)$$

Let $x_1(t) = y(t)$. Then an equivalent set of first order equations written in vector form is

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} -a_2 & 1 & 0 \\ -a_1 & 0 & 1 \\ -a_o & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} b_2 \\ b_1 \\ b_o \end{bmatrix} u(t)$$

The linear version of the discrete-time system (1.2) is

$$\underline{x}(k+1) = \underline{F}(k)\underline{x}(k) + \underline{G}(k)\underline{u}(k) \tag{1.4}$$

The solution of the discrete-time linear system takes the following form;

$$\underline{x}(n) = \underline{\Phi}(n,k)\underline{x}(k) + \sum_{j=k}^{n-1} \underline{\Phi}(n,j+1)\underline{G}(j)\underline{u}(j) \qquad , \qquad n > k$$

where

$$\underline{\Phi}(n,k) = \underline{F}(n-1)\, \underline{\Phi}(n-1,k) = F(n-1) \cdots F(k) \qquad , \qquad n > k$$

$$\underline{\Phi}(k,k) = \underline{I}$$

Procedures for reducing difference equations to a set of first order difference equations are analogous with the ones illustrated for differential equations. For a more detailed discussion of linear systems from the state point of view see [33].

1.4 Mathematical Model

1.41 Statement of the Continuous-Time Problem

We begin our discussion with a mathematical statement of the continuous-time estimation problem. The motivation for the mathematical model used will become apparent in the discussion immediately following. The reader may wish to read over this section lightly at first, returning after the relation of the model to various problems is developed.

The problem to be considered is the estimation of the state variables of a nonlinear dynamic system which is excited by white Gaussian noise. It is assumed that a nonlinear combination of the state variables corrupted by additive white Gaussian noise may be observed over a finite time interval.

In particular, we consider processes which may be described by a

system of nonlinear differential equations which can be written in vector notation as

$$\dot{\underline{x}} = \underline{f}(\underline{x}(t),t) + \underline{G}(t)\underline{w}(t) \tag{1.5}$$

The observed signal is

$$\underline{z}(t) = \underline{h}(\underline{x}(t),t) + \underline{v}(t) \tag{1.6}$$

where;

$\underline{x}$ is an n-vector, the state vector of the system

$\underline{w}$ is an m-vector, $m \leq n$, the random input

$\underline{z}$ is a p-vector, the observation

$\underline{G}$ is an n x m matrix

$\underline{v}$ is a p-vector, the measurement noise

The functions $\underline{w}$ and $\underline{v}$ are white Gaussian noise processes with zero means and the following covariance matrices;

$$\left.\begin{aligned} E\left[\underline{w}(t)\underline{w}'(\tau)\right] &= \underline{Q}(t) \quad \delta(t-\tau) \\ E\left[\underline{v}(t)\underline{v}'(\tau)\right] &= \underline{R}(t) \quad \delta(t-\tau) \\ E\left[\underline{w}(t)\underline{v}'(\tau)\right] &= \underline{0} \end{aligned}\right\} \quad \text{for all } t \text{ and } \tau$$

where E is the expectation operator, δ is the Dirac delta function, and $\underline{Q}(t)$ and $\underline{R}(t)$ are positive definite symmetric matrices which are continuously differentiable in t. It will be necessary to assume that $\underline{f}$ and $\underline{h}$ satisfy certain differentiability conditions with respect to their arguments.

This model may be represented by the block diagram of Fig. 1. The wide arrows are used to indicate the flow of vector quantities.

The output of an oval block is obtained by the indicated nonlinear operation on its input.

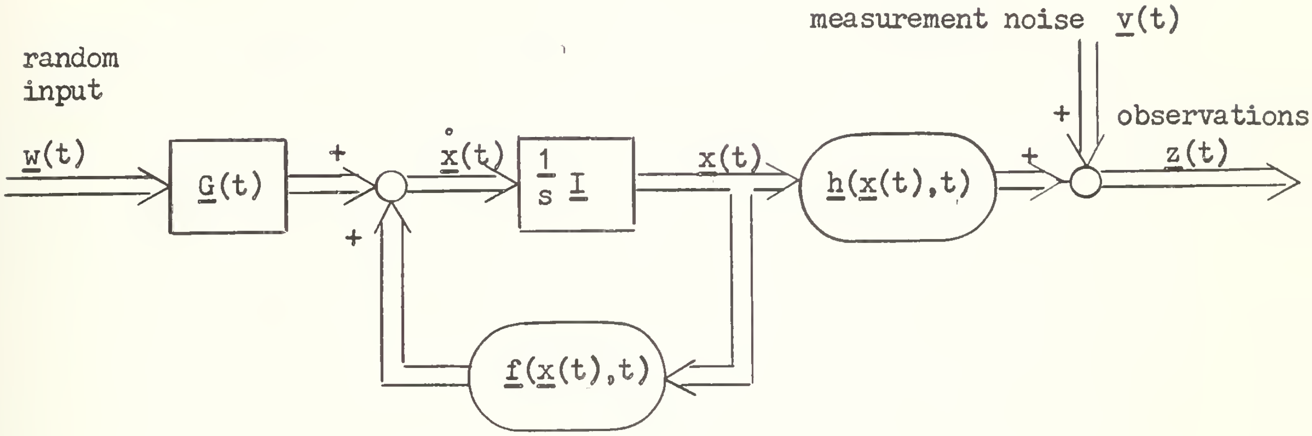


Fig. 1. Model of the Continuous-Time System

Observations begin at an initial time t_o. The problem of real time estimation of state variables is to use the history of observations from t_o up to the present time t (i.e., $\underline{z}_{[t_o, t]}$) to continually produce up-to-date estimates of the present state $\underline{x}(t)$.

It will be convenient to imbed this problem in the more general problem of estimating the entire segment $\underline{x}_{[t_o, t+T]}$. This point will be discussed in subsequent chapters.

1.42 Relation to Optimum Control Problems

The so-called optimum control problem is usually formulated in the following manner. The physical process to be controlled is represented by a system of ordinary differential equations of the form

$$dx_i(t)/dt = f_i(x_1, \ldots, x_n, u_1, \ldots, u_r, t) \qquad , i = 1, \ldots, n$$

where $x_1, \ldots, x_n$ are the state variables of the process, $u_1, \ldots, u_r$

are the control variables and t is time. These equations are usually written in vector notation as

$$\underline{\mathring{x}}(t) = \underline{f}(\underline{x}(t), \underline{u}(t), t)$$

The control problem is that of choosing $\underline{u}(t)$ from a set of admissible controls so that the state of the system follows an optimal trajectory. Optimality is usually defined as the minimization of a scalar performance index which may in general depend on $\underline{u}$, $\underline{x}$ and t. For example, if it were desired to move the system from an initial state $\underline{x}_0$ to a specified final state $\underline{x}_1$, the performance index might include time of transit, energy expended, or a combination of these and other criteria.

The practical considerations of model inaccuracies and external disturbances make it desirable to have a closed loop system, that is, one in which the control $\underline{u}(t)$ is specified as a function of the state $\underline{x}(t)$ rather than as a preprogrammed function of time.

Despite its short history, the literature devoted to various aspects of this problem is voluminous. Various mathematical techniques including the calculus of variations [14,18,36] , the "Maximum Principle" of Pontryagin [52,55,41], Bellman's dynamic programming [3,5,16] , and steepest descent [9,35] have been used in connection with this class of problems. In almost all of this work it is necessary to assume that the state variables of the system are observable or exactly measurable or may be calculated instantaneously from other observable quantities. Because of external disturbances, instrument noise and the intrinsic difficulty of physically performing the unbounded operation of differentiation, the assumption that the state variables of the system are known exactly

is almost never justified.

It is known that for linear systems with quadratic performance criteria it is possible to solve the estimation problem and the optimization problem separately and still obtain the over-all optimum system (see [23] or [28]). No analogous result can be anticipated for nonlinear systems; that is, when estimates are used in place of the actual values in nonlinear systems we cannot assume that the over-all system will still be optimal. From the viewpoint of application, however, if estimates of state variables were available, one would have no alternative but to use them, since the problem of joint estimation and optimization for a nonlinear system belongs to a class of extremely difficult unsolved statistical optimization problems. Indeed, either the estimation problem or the optimization problem alone is extremely formidable.

A diagram of the over-all control system using the estimates of the state variables is shown in Fig. 2. The estimator operates on all observable quantities, in this case $\underline{u}$ and $\underline{z}$, to produce an estimate of

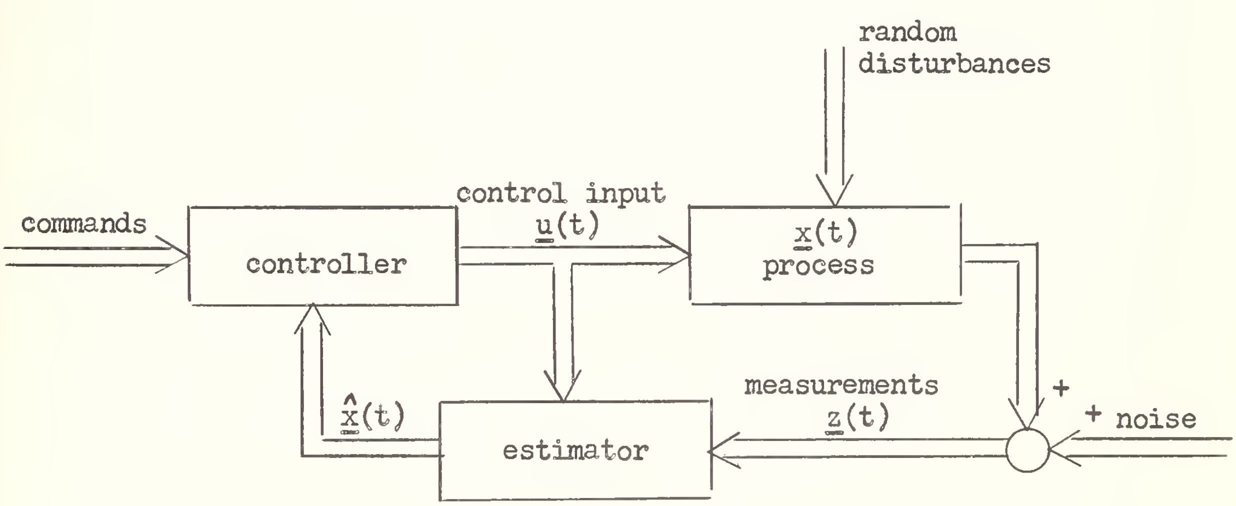


Fig. 2. Block diagram of over-all control system

$\underline{x}$ which it feeds to the controller. The controller produces $\underline{u}$ which acts on the process. The value of $\underline{u}$ is also fed to the estimator.

The mathematical model for this noisy system is

$$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t),\underline{u}(t),t)+\underline{G}(t)\underline{w}(t)$$

$$\underline{z}(t) = \underline{h}(\underline{x}(t),t) + \underline{v}(t)$$

This is identical with the model of (1.5) and (1.6) except that we have the additional known quantity $\underline{u}$ appearing in the first equation. In our model this is accounted for by the explicit dependence of $\underline{f}(\underline{x}(t),t)$ in (1.5) on the time parameter t.

There is no loss of generality in assuming $\underline{w}(t)$ to be white noise. It has been emphasized in the early work of Bode and Shannon [8], and Zadeh and Ragazzini [67] that a Gaussian process with a rational spectrum may be considered equivalent to the output of a linear system excited by white Gaussian noise. If we denote the state variables of this equivalent linear system by $\underline{x}^{(2)}$ and the state variables of the process to be controlled by $\underline{x}^{(1)}$, we may form a composite state vector

$$\underline{x} = \begin{bmatrix} \underline{x}^{(1)} \\ \underline{x}^{(2)} \end{bmatrix}$$

and obtain equations of the form of (1.5) and (1.6).

A simple example may help to clarify this point. Consider a scalar process of the form

$$\dot{x}(t) = f(x(t),u(t)) + r(t)$$

$$z(t) = h(x(t),t) + v(t)$$

where r is a Gaussian random process with the spectrum

$$\Phi_{rr}(\omega) = \frac{N^2}{\omega^2 + a^2}$$

Then we may consider r as the output of the linear system

$$\dot{r} = - a\, r + w$$

where w is a white Gaussian noise process with the spectrum

$$\Phi_{ww}(\omega) = N^2$$

If we designate x by $x^{(1)}$ and r by $x^{(2)}$, we obtain the following system of equations;

$$\frac{d}{dt}\begin{bmatrix} x^{(1)}(t) \\ x^{(2)}(t) \end{bmatrix} = \begin{bmatrix} f(x^{(1)}(t),u(t)) + x^{(2)}(t) \\ - ax^{(2)}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t)$$

$$z(t) = h(x^{(1)}(t),t) + v(t)$$

which are of the form of (1.5) and (1.6).

1.43 Relation to Adaptive Control

Although there is no universally accepted definition of adaptive control, most investigators would agree that a central problem in the theory of adaptive control is the identification problem [46]. Changing environmental conditions are frequently incorporated in the mathematical model of the system as changes of certain parameters in that model. The

identification problem is concerned with automatically evaluating current values of the parameters. Because these parameters usually enter the model multiplicatively, the problem of estimating them is often equivalent to that of estimating state variables of a nonlinear dynamic system.

Estimation of parameters will be treated as a special case of estimation of state variables. The basic technique is to augment the state-space of the system to include the parameters themselves and additional quantities to account for the nature of the statistical variations of the parameters. For example, an unknown constant parameter b would be handled by letting x_{n+1} equal b and adjoining the equation $\dot{x}_{n+1}(t) = 0$ to the basic system of equations describing the process. A randomly varying parameter is treated as the output of a linear system excited by white noise. In this case it is necessary to adjoin the states of this fictitious linear system to the states of the basic process.

Let us consider a simple example to illustrate these ideas. A problem of considerable importance is the control of systems with lightly damped mechanical resonances [46]. Frequently there is a problem of tuning the compensating network as the resonant frequency of the system drifts due to changing environmental conditions, such as are commonly experienced by high performance aircraft. For a second order system a suitable model for this situation might be

$$\ddot{y} + b\,\dot{y} + (c_o + c)\,y = u + w_1$$

$$\dot{c} + a\,c = w_2$$

$$z = y + v$$

where the random parameter c is considered to be the output of a first order system excited by white noise. Letting

$x_1 = y \qquad x_2 = \dot{y} \qquad x_3 = c$

we may rewrite these equations in the form of (1.5) and (1.6) as follows;

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_2 \\ -b\, x_2 - (c_o + x_3)\, x_1 + u \\ -a\, x_3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$z = x_1 + v$$

Again the explicit dependence of $\underline{f}(\underline{x}(t),t)$ in (1.5) on t accounts for any known inputs such as control forces or test signals.

1.44 Relation to Linear Filtering Theory

The application of statistical methods to automatic control problems began with the classic work of Wiener [65] on linear filtering and prediction theory. The early exposition by Bode and Shannon [8] presented the principal results of the Wiener theory in a form more easily understood by engineers. At about the same time, an extension of the Wiener theory which was particularly significant from a control point of view was made by Zadeh and Ragazzini [67]. Both of these early papers [8, 67] emphasized the viewpoint that stationary random processes with rational spectra could be thought of as the output of linear systems excited by white noise. Several investigators [1,13,27] have extended these ideas to multidimensional systems. The work of Davis [13] gives a general solution to the problem of factoring spectral matrices and relates the linear stationary theory to the state-space approach. Today material on the one-dimensional Wiener theory and its extensions is available in a large number of standard textbooks; for example, [11,12,39,42,

49, 57, 60, 63].

The problem we consider is a nonlinear version of the growing memory filters discussed by Kalman and Bucy [34] who studied random processes described by the linear vector differential equations

$$\overset{\circ}{\underline{x}}(t) = \underline{F}(t)\underline{x}(t) + \underline{G}(t)\underline{w}(t) \tag{1.7}$$

$$\underline{z}(t) = \underline{H}(t)\underline{x}(t) + \underline{v}(t) \tag{1.8}$$

which are the linear form of (1.5) and (1.6). They give the general solution to the problem of linearly estimating and predicting $\underline{x}(t)$ based on a finite observation interval. If $\underline{F}$, $\underline{G}$, and $\underline{H}$ are constant matrices and the observation interval extends to negative infinity, this is a multivariate Wiener filtering problem in which the random process is specified by the linear system (1.7) and (1.8) rather than by its spectral matrix. For this stationary case, (1.7) and (1.8) may be obtained by factorization of the spectral matrix [13]. It is indeed unfortunate that the important results of Kalman and Bucy are not well understood by many control specialists. This is, no doubt, due in part to the elegant but intricate nature of their original derivation. An interesting by-product of our study of the nonlinear problem has been a simplified derivation of their results for the corresponding linear problem. (see Chapter III.)

The idea of considering a random process to be the result of exciting a linear system with white noise has great intuitive appeal. This is especially true for a stationary process where a suitable linear system may be postulated as the result of direct measurement of the spectrum of the process. For the time-varying system of Kalman and Bucy it is not at all clear how to make measurements on a non-stationary process to find a suitable system. However, their work has direct application to control problems

in which prior information about a time-varying system is available.

There is considerable appeal to the idea of considering a stationary non-Gaussian random process to be the result of exciting with white noise a nonlinear time-invariant system having a finite number of state variables. Again it is not at all clear how to make measurements on the process to determine the appropriate system. Our work is applicable to cases in which prior information about the nonlinear system is available.

1.45 Generality of the Model

It has already been pointed out that any known inputs such as control forces or test signals are taken into account by the explicit dependence of $\underline{f}(\underline{x}(t),t)$ on t. We also showed how estimation of parameters could be treated as a special case of estimation of state variables.

There is no loss of generality in assuming $\underline{Q}(t)$ is positive definite because of the presence of $\underline{G}(t)$ in (1.5). The restriction that $\underline{R}(t)$ be positive definite is more basic. To relax this restriction for the continuous-time problem is to admit the possibility of differentiating the observed signal [29].

A more general version of (1.5) would be

$$\overset{\circ}{\underline{x}}(t) = \underline{f}(\underline{x}(t),\underline{w}(t),t) \tag{1.9}$$

The basic objection to (1.9) is that extreme care must be taken when performing nonlinear operations on white noise because white noise is defined by a limiting process and the appropriate limit may fail to exist. For example, it is meaningless to square white noise. This objection may be avoided by considering the equation

$$\overset{\circ}{\underline{x}}(t) = \underline{f}(\underline{x}(t),t) + \underline{G}(\underline{x}(t),t)\underline{w}(t) \tag{1.10}$$

For simplicity, we shall, in the main, confine our attention to systems in which the white noise enters linearly as in (1.5), pointing out where extensions can be made to include systems of the type (1.10) and well defined versions of (1.9).

1.5 An Intuitive Approach

Before becoming involved in a maze of equations it is perhaps worthwhile to see in which direction an intuitive approach might lead.

Consider the problem of estimating the state variables of a nonlinear system for which all inputs are known (no random disturbances) and the observations at the output are noisy. A simple approach to this problem is to feed the known inputs to a model of the nonlinear system which hopefully, after some initial transients have subsided, will behave in the same manner as the system itself. The model would be constructed so that its state variables were easily measurable. Such a scheme is shown in Fig. 3.

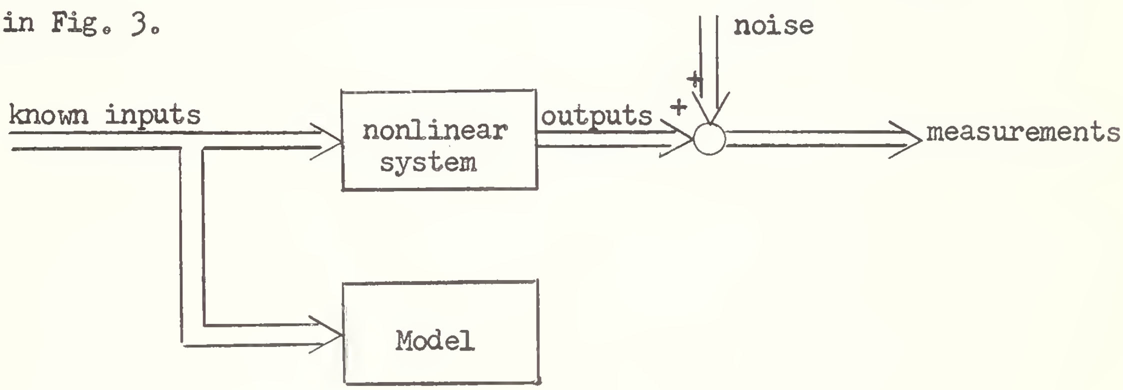


Fig. 3. Open loop model approach

We shall say that the scheme of Fig. 3 uses the model in an open loop manner. While in theory this approach should yield good results, the practical considerations of model inaccuracies and random disturbances would render it useless in all but the simplest of cases.

The obvious way to combat model inaccuracies and random disturbances is by the use of feedback. This is done in Fig. 4.

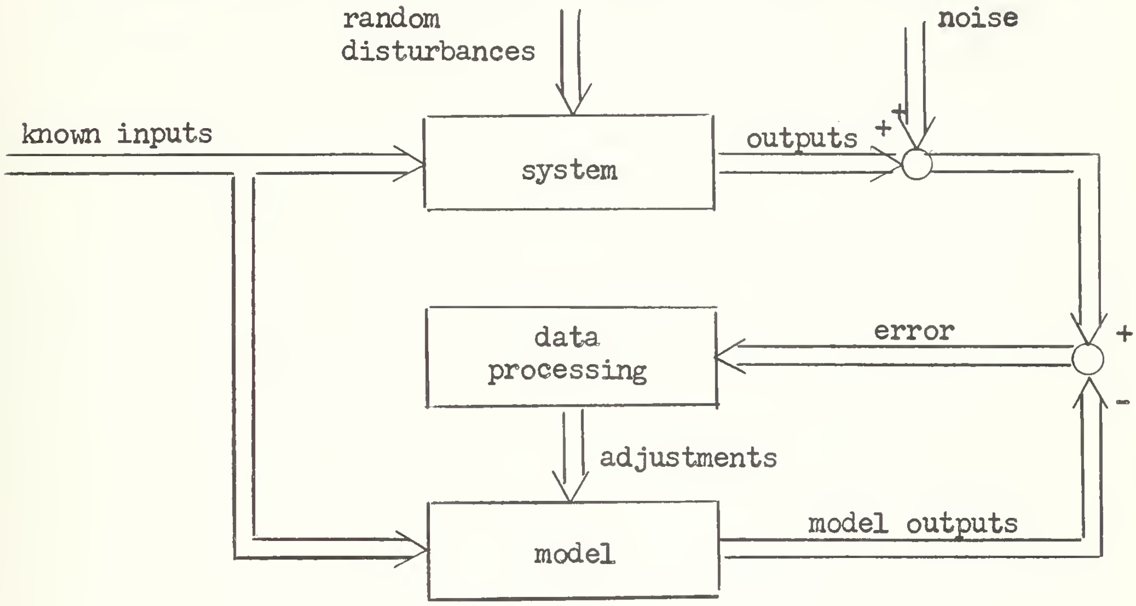


Fig. 4. Closed loop model approach

The closed loop configuration of Fig. 4 has great intuitive appeal. While it is not clear how the error should be processed in order to properly adjust the model, we would expect that something could be worked out. It will be shown later that the model approach may be derived as the result of considering the basic equations (1.5) and (1.6) and that there is a rational way of processing the data.

It is amazing how many seemingly diversified topics fall into the class of closed loop model schemes. Kalman and Bucy [34] show that the solution of the linear filtering problem may be interpretated in this manner. Davis [13] gives a model interpretation when the measurement noise is not white. The very general Wiener theory of nonlinear systems [46, 66] assumes that nothing is known about the system and, using white Gaussian noise as the known test input, develops a model in terms of an orthogonal series.

A variant on this theme is the optimization of the system by adjusting it to behave more like the model. The recent work of Sakrison [56] and Kushner [38] in this area may be cited as examples. A large number of optimization and/or learning schemes using models are to be found in the adaptive control literature [43, 46].

The various schemes using models differ in the amount of prior information assumed and the way in which observations are processed to make improvements on the model or system. In our work it is assumed that the system is represented by known equations of the type (1.5) and (1.6). In the preceding sections it was shown that many problems fall into this category. There are many situations, however, in which adequate knowledge of the physics of the system is not available and one is unable to write equations of the form of (1.5) and (1.6) even using the technique of treating unknown parameters as additional state variables. Our work is not directly applicable to these situations.

CHAPTER II

ESTIMATION OF STATE VARIABLES FOR DISCRETE-TIME SYSTEMS

2.1 Introduction

This chapter is devoted to the problem of estimating state variables for discrete-time systems. This problem is completely analogous to the continuous-time problem discussed in the preceding chapter. The discussion of the usefulness and limitations of the continuous-time model presented there applies also to the discrete-time model to be introduced forthwith.

Discrete-time systems are usually approximations of continuous-time systems. They assume great practical importance because digital computers work in discrete-time and complicated problems usually must be solved on digital computers. The extent to which the ability to use digital computers to obtain numerical solutions to complex problems has influenced modern technology is evidenced in the fact that today an effective computational algorithm is often considered to be a solution to a problem which, because of the limitations of purely analytical techniques, would not even have been investigated a decade ago. The possibility of using special purpose digital computers as control system components is to a large extent responsible for present trends in control theory. We shall continually be mindful of computational difficulties as the investigation of the estimation problem proceeds.

The equations which characterize the remainder of this work seem quite formidable at first glance. The reader should take heart in that the underlying ideas are simple and most derivations involve nothing more complicated than differentiation.

2.2 Problem Statement

The problem to be considered is the estimation of the state variables of a discrete-time nonlinear system which is excited by a sequence of independent Gaussian random vectors. It is assumed that nonlinear combinations of the state variables corrupted by additive independent Gaussian noise may be observed.

In particular, we consider systems which are described by vector difference equations of the following form;

$$\underline{x}(k+1) = \underline{f}(\underline{x}(k), k) + \underline{G}(k)\underline{w}(k) \tag{2.1}$$

The observed signal is

$$\underline{z}(k) = \underline{h}(\underline{x}(k), k) + \underline{v}(k) \tag{2.2}$$

where

$\underline{x}$ is an n-vector, the state vector of the system

$\underline{w}$ is an m-vector, $m \leq n$, the random input

$\underline{z}$ is a p-vector, the observation

$\underline{G}$ is a n x m matrix

$\underline{v}$ is a p-vector, the measurement noise

$\underline{h}$ and $\underline{f}$ are vector-valued functions

$\ldots, \underline{w}(k), \underline{w}(k+1), \ldots$ and $\ldots, \underline{v}(k), \underline{v}(k+1), \ldots$

are sequences of Gaussian random vectors with zero means and the following covariance matrices;

$$\left.\begin{aligned} E[\underline{w}(j)\ \underline{w}'(k)] &= \underline{Q}(k)\ \delta_{jk} \\ E[\underline{v}(j)\ \underline{v}'(k)] &= \underline{R}(k)\ \delta_{jk} \\ E[\underline{w}(j)\ \underline{v}'(k)] &= \underline{0} \end{aligned}\right\} \text{ for all intergers j and k}$$

$\underline{Q}$ and $\underline{R}$ are positive definite matrices. E is the expectation operator and δ is the Kronecker delta. Again there is no loss of generality by assuming that $\underline{Q}$ is positive definite. The restriction that $\underline{R}$ be positive definite will later be relaxed.

Having observed a finite sequence $\{\underline{z}(o),\ldots,\underline{z}(n)\}$, we may, in general, seek an estimate of an entire sequence of states $\{\underline{x}(o),\ldots,\underline{x}(n),\ldots\underline{x}(n+m)\}$. We shall call this the estimation problem. In this formulation we include as special cases the filtering problem, where an estimate of the current state $\underline{x}(n)$ is sought; the smoothing problem, where an estimate of the sequence $\{\underline{x}(o),\ldots,\underline{x}(n)\}$ is of interest; and the prediction problem, where an estimate of a future state $\underline{x}(n+m)$ is desired.

As in the continuous-time problem, we may easily include as state variables unknown or randomly varying parameters. Any known forcing functions, such as control inputs or test signals are accounted for by the explicit dependence of $\underline{f}(\underline{x}(k),k)$ on the time parameter k. Such known parameters will usually simplify the estimation problem since they provide additional information about the system behavior.

The model for the discrete-time system may be represented by the block diagram of Fig. 5.

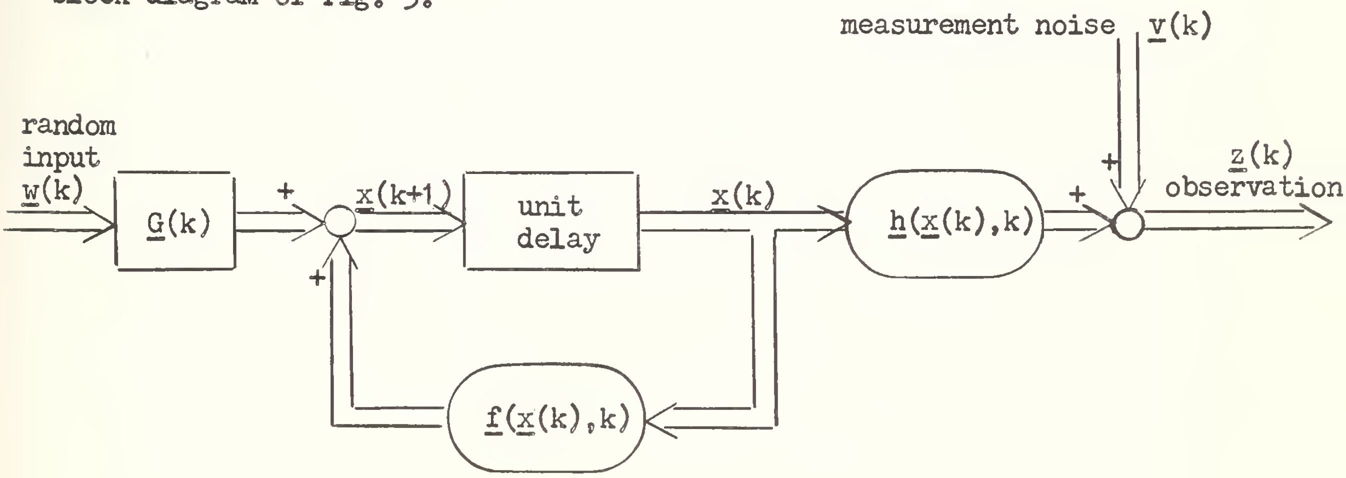


Fig. 5. Block diagram of the discrete-time system

Kalman [31] first solved the linear filtering and prediction problems for the linear version of (2.1) and (2.2) given by

$$\underline{x}(k+1) = \underline{F}(k)\underline{x}(k) + \underline{G}(k)\underline{w}(k)$$

$$\underline{z}(k) = \underline{H}(k)\underline{x}(k) + \underline{v}(k)$$

In recent independent work Ho [25] has used an approach similar to the one we shall use to re-solve this linear problem.

The systems described by (2.1) are a subset of the systems defined by the more general equation

$$\underline{x}(k+1) = \underline{f}(\underline{x}(k),\underline{w}(k),k) \tag{2.3}$$

Extensions to include these more general systems will be pointed out, but it is more convenient to work with (2.1) as the basic equation.

Finally, we should note that the time parameter k designates the time of the kth observation. There is no need to assume that observations are equally spaced. There is no loss of generality in assuming that observations begin at k=o.

2.3 The Probability Distribution for the Sequence of States

Let us first examine the problem of estimating the sequence of states $\{\underline{x}(o),\ldots,\underline{x}(n)\}$, having observed the sequence $\{\underline{z}(o),\ldots,\underline{z}(n)\}$. Proceeding in a straightforward manner, we consider the *a posteriori* probability density function

$$p_{x_N|z_N}(\underline{x}(o),\ldots,\underline{x}(n)|\underline{z}(o),\ldots,\underline{z}(n))$$

which is the conditional joint probability density function for the

sequence of states $\{\underline{x}(o),\ldots,\underline{x}(n)\}$ on the assumption that the observation sequence $\{\underline{z}(o),\ldots,\underline{z}(n)\}$ has occurred. This function will be, in general, extremely complex. Using Bayes' Theorem, we obtain

$$p_{x_N|z_N}(\underline{x}(o),\ldots,\underline{x}(n)|\underline{z}(o),\ldots,\underline{z}(n))$$

$$= \frac{p_{z_N|x_N}(\underline{z}(o),\ldots,\underline{z}(n)|\underline{x}(o),\ldots,\underline{x}(n))\; p_{x_N}(\underline{x}(o),\ldots,\underline{x}(n))}{p_{z_N}(\underline{z}(o),\ldots,\underline{z}(n))} \tag{2.4}$$

Let us consider the terms of interest which appear in the numerator of this expression. Using (2.2) and the independence of the measurement noise, we may rewrite the first term as

$$p_{z_N|x_N}(\underline{z}(o),\ldots,\underline{z}(n)|\underline{x}(o),\ldots,\underline{x}(n)) = \prod_{k=o}^{n} p_{v_k}(\underline{z}(k)-\underline{h}(\underline{x}(k),k))$$

where p_{v_k} is the Gaussian density function for the additive measurement noise $\underline{v}(k)$.

The second term may be rewritten in the following form;

$$p_{x_N}(\underline{x}(o),\ldots,\underline{x}(n)) = p_o(\underline{x}(o)) \cdot p_1(\underline{x}(1)|\underline{x}(o)) \cdots$$

$$p_n(\underline{x}(n)|\underline{x}(n-1),\ldots,\underline{x}(o))$$

The independence of successive values of $\underline{w}$ makes (2.1) a Markov process [26]. That is,

$$p_k(\underline{x}(k)|\underline{x}(k-1),\ldots,\underline{x}(o)) = p_k(\underline{x}(k)|\underline{x}(k-1))$$

Substituting into (2.4), we obtain

$$p_{x_N|z_N}(\underline{x}(o),\ldots,\underline{x}(n)|\,\underline{z}(o),\ldots,\underline{z}(n))$$

$$= \frac{\prod_{k=o}^{n} p_{v_k}(\underline{z}(k)-\underline{h}(\underline{x}(k),k))\; p_o(\underline{x}(o)) \;\prod_{k=1}^{n} p_k(\underline{x}(k)|\underline{x}(k-1))}{p_{z_N}(\underline{z}(o),\ldots,\underline{z}(n))} \tag{2.5}$$

We assign an a priori Gaussian distribution with mean $\underline{m}$ and covariance matrix $\underline{P}(o)$ for the unknown initial state $\underline{x}(o)$. It will be convenient, although not essential, to assume $\underline{P}(o)$ is non-singular. This restriction will later be relaxed.

Before proceeding further, it is necessary to examine the nature of the input sequence. Let $\underline{r}(k) = \underline{G}(k)\underline{w}(k)$. Then $\ldots,\underline{r}(k),\underline{r}(k+1),\ldots$ is a sequence of independent Gaussian random vectors with zero means. The covariance matrix for $\underline{r}(k)$ is $\underline{G}(k)\underline{Q}(k)\underline{G}'(k)$. If we designate the probability density function for $\underline{r}(k)$ by p_{r_k} and then use the fact that

$$\underline{x}(k) = \underline{f}(\underline{x}(k-1),k-1) + \underline{r}(k-1)$$

we obtain formally

$$p_k(\underline{x}(k)|\underline{x}(k-1)) = p_{r_{k-1}}(\underline{x}(k)-\underline{f}(\underline{x}(k-1),k-1))$$

If $\underline{GQG}'$ is singular, then $\underline{r}$ is confined with probability one to a hyperplane in n-space of dimension equal to the rank of $\underline{GQG}'$, and without resorting to delta functions we are unable to write an explicit expression for the probability density function p_{r_k}. We note, however, formally

$$p_{r_k}(\underline{r}(k)) = \int p_{r_k|w_k}(\underline{r}(k)|\,\underline{w}(k))\; p_{w_k}(\underline{w}(k))\; d\underline{w}(k) \tag{2.6}$$

where the conditional density function $p_{r_k|w_k}$ may be interpreted as a constraint, since $\underline{r}(k)$ is known once $\underline{w}(k)$ is specified.

If, on the other hand, the covariance matrix $\underline{GQG}'$ is non-singular, an explicit expression may be given for the Gaussian probability density function p_{r_k}. In this case all the terms in the numerator of (2.5) may be expressed explicitly since p_{r_k}, p_v, and p_o are all Gaussian with known means and covariance matrices. Then, (2.5) becomes

$$p_{x_N|z_N}(\underline{x}(o),\ldots,\underline{x}(n)|\underline{z}(o),\ldots\underline{z}(n)) = C(\underline{z}(o),\ldots,\underline{z}(n)) \cdot$$

$$\exp\left\{-\tfrac{1}{2}\sum_{k=0}^{n}\|\underline{z}(k)-\underline{h}(\underline{x}(k),k)\|^2_{\underline{R}^{-1}(k)} - \tfrac{1}{2}\|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)} - \tfrac{1}{2}\sum_{k=0}^{n-1}\|\underline{x}(k+1)-\underline{f}(\underline{x}(k),k)\|^2_{[\underline{GQG}']^{-1}}\right\} \qquad (2.7)$$

where $C(\underline{z}(o),\ldots,\underline{z}(n))$ is a normalizing factor depending only on the observed output sequence and known constants. The norm notation has been used to emphasize the positiveness of the quadratic forms appearing in this expression.

2.4 The Estimation Procedure

Now that an expression for the *a posteriori* probability density function has been obtained for the case in which $\underline{GQG}'$ is non-singular, the nature of the difficulties introduced by nonlinearities is apparent. One might choose an estimate $\hat{\underline{x}}$ so as to minimize the conditional expectation of some appropriate loss function. That is

$$\underset{\hat{\underline{x}}}{\mathrm{Min}}\left\{\underset{x|z}{E}\left[L(\underline{x}-\hat{\underline{x}})\right]\right\}$$

For a quadratic loss function defined by a positive definite matrix $\underline{W}$, it is well known [29] that the best estimate is the conditional mean. To see this let

$$L_1(\underline{x}-\hat{\underline{x}}) = \|\underline{x}-\hat{\underline{x}}\|^2_{\underline{W}} = \underline{x}'\underline{W}\underline{x} - 2\hat{\underline{x}}'\underline{W}\underline{x} + \hat{\underline{x}}'\underline{W}\hat{\underline{x}}$$

Then

$$\underset{x|z}{E}\left[L_1(\underline{x}-\hat{\underline{x}})\right] = \overline{\underline{x}'\underline{W}\underline{x}} - 2\hat{\underline{x}}'\underline{W}\overline{\underline{x}} + \hat{\underline{x}}'\underline{W}\hat{\underline{x}}$$

Differentiating with respect to $\hat{\underline{x}}$ and setting the result equal to zero we find that the best estimate is $\overline{\underline{x}}$, the conditional mean.

The task of finding the mean of (2.7) is overwhelming because of the nonlinearity of the system.

It is also well known [54] that for a linear loss function of the form

$$L_2(\underline{x}-\hat{\underline{x}}) = \sum_i c_i \,|x_i-\hat{x}_i| \qquad\qquad c_i > 0$$

the ith coordinate of the best estimate is the median of the marginal distribution for x_i. To see this note that

$$\underset{x|z}{E}\left[L_2(\underline{x}-\hat{\underline{x}})\right] = \sum_i c_i \,\overline{|x_i-\hat{x}_i|}$$

and

$$\overline{|x_i-\hat{x}_i|} = \int_{-\infty}^{\hat{x}_i} (\hat{x}_i-x_i)\, p_i(x_i)\, dx_i + \int_{\hat{x}_i}^{\infty} (x_i-\hat{x}_i)\, p_i(x_i)\, dx_i$$

Taking the derivative of this expression with respect to $\hat{x}_i$ and setting

the result equal to zero, we obtain the desired result,

$$\int_{-\infty}^{\hat{x}_i} p_i(x_i)\, dx_i = \int_{\hat{x}_i}^{\infty} p_i(x_i)\, dx_i$$

Again the nonlinearity of the system is our nemisis if we try to apply this result to (2.7).

An appealing loss function for many applications is one that weights equally all errors greater than some threshold value. This would express the idea that once the size of the error exceeded a specified level, an increase in the size of the error would not be significant. An example of such a loss function is

$$L_3(\underline{x}-\underline{\hat{x}}) = \begin{cases} \| \underline{x}-\underline{\hat{x}} \|^2_{\underline{W}} & , \text{ for } \| \underline{x}-\underline{\hat{x}} \|^2_{\underline{W}} \leq c \\ c & , \text{ for } \| \underline{x}-\underline{\hat{x}} \|^2_{\underline{W}} > c \end{cases}$$

For this type of loss function we would expect that the mode would be a better estimate than either the mean or the median. To gain some insight into this situation, consider the loss function and the probability density function for a scalar random variable θ shown in Fig. 6.

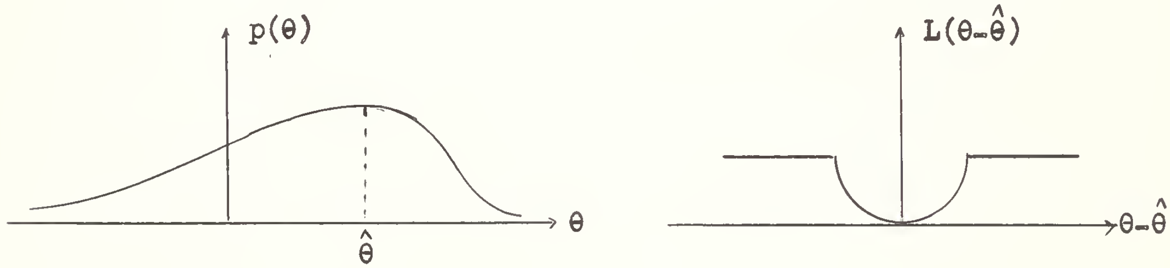


Fig. 6. A probability density function and a loss function

It is clear that for the situation of Fig. 6 the expected loss

(area of the product) will be minimized by locating the valley of $L(\theta-\hat{\theta})$ to correspond to the peak of $p(\theta)$. That is, the mode would be the suitable estimate.

It is known (see [29] or [58]) that if the distribution function $P_{x|z}(x)$ is unimodal and symmetric about its mean, then the conditional mean is the best estimate for a large class of symmetric loss functions including L_1, L_2 and L_3 above. We hasten to point out that in this case the mean and the mode coincide and one might just as well say that the mode were the best estimate.

The estimate we shall use is the mode of the a posteriori distribution. This is closely related to the method of maximum likelihood; the difference being that we take the Bayesian point of view of assigning an a priori distribution to the initial state.

For $\underline{GQG}'$ non-singular, the estimate will be obtained by minimizing

$$2J_n = \| \underline{x}(o)-\underline{m} \|^2_{\underline{P}^{-1}(o)} + \sum_{k=o}^{n} \| \underline{z}(k)-\underline{h}(\underline{x}(k),k) \|^2_{\underline{R}^{-1}(k)}$$

$$+ \sum_{k=o}^{n-1} \| \underline{x}(k+1)-\underline{f}(\underline{x}(k),k) \|^2_{[\underline{GQG}']^{-1}} \tag{2.8}$$

with respect to the sequence $\{\underline{x}(o),\ldots,\underline{x}(n)\}$. This is equivalent to minimizing with respect to the sequences $\{\underline{x}(o),\ldots,\underline{x}(n)\}$ and $\{\underline{w}(o),\ldots,\underline{w}(n-1)\}$, the expression

$$\|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)} + \sum_{k=o}^{n} \|\underline{z}(k)-\underline{h}(\underline{x}(k),k)\|^2_{\underline{R}^{-1}(k)} + \sum_{k=o}^{n-1} \|\underline{w}(k)\|^2_{\underline{Q}^{-1}(k)}$$

subject to the constraint

$$\underline{x}(k+1) = \underline{f}(\underline{x}(k),k) + \underline{G}(k)\underline{w}(k)$$

Introducing an n-vector of Lagrange multipliers $\underline{\lambda}$ to incorporate the constraint, we may minimize

$$I_n = \tfrac{1}{2}\|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)} + \sum_{k=o}^{n} \tfrac{1}{2}\|\underline{z}(k)-\underline{h}(\underline{x}(k),k)\|^2_{\underline{R}^{-1}(k)}$$

$$+ \sum_{k=o}^{n-1}\left\{\tfrac{1}{2}\|\underline{w}(k)\|^2_{\underline{Q}^{-1}(k)} + \underline{\lambda}'(k)\left[\underline{x}(k+1)-\underline{f}(\underline{x}(k),k)-\underline{G}(k)\underline{w}(k)\right]\right\} \quad (2.9)$$

From (2.6) we see that the function I_n may also be minimized when $\underline{GQG}'$ is singular.

The more general problem of estimating the state sequence $\{\underline{x}(o),\ldots,\underline{x}(n),\ldots,\underline{x}(n+m)\}$ after the sequence $\{\underline{z}(o),\ldots,\underline{z}(n)\}$ has been observed may be reduced by an analogous development to that of minimizing

$$I_{n,m} = \tfrac{1}{2}\|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)} + \sum_{k=o}^{n} \tfrac{1}{2}\|\underline{z}(k)-\underline{h}(\underline{x}(k),k)\|^2_{\underline{R}^{-1}(k)}$$

$$+ \sum_{k=o}^{n+m-1}\left\{\tfrac{1}{2}\|\underline{w}(k)\|^2_{\underline{Q}^{-1}(k)} + \underline{\lambda}'(k)\left[\underline{x}(k+1)-\underline{f}(\underline{x}(k),k)-\underline{G}(k)\underline{w}(k)\right]\right\} \quad (2.10)$$

A comparison of (2.9) and (2.10) yields

$$I_{n,m} = I_n + \sum_{k=n}^{n+m-1}\left\{\tfrac{1}{2}\|\underline{w}(k)\|^2_{\underline{Q}^{-1}(k)} + \underline{\lambda}'(k)\left[\underline{x}(k+1)-\underline{f}(\underline{x}(k),k)-\underline{G}(k)\underline{w}(k)\right]\right\} \quad (2.11)$$

From (2.11) we conclude that $I_{n,m}$ may be minimized by first minimizing I_n and then setting $\underline{w}(k)$ equal to zero for $k=n,\ldots,n+m-1$. This leads us to the important conclusion that predictions are simply extrapolations of present estimates. Specifically, if we let $\hat{\underline{x}}(k|n)$ denote the estimate of $\underline{x}(k)$ given $\{\underline{z}(o),\ldots,\underline{z}(n)\}$, then

$$\hat{\underline{x}}(n+1|n) = \underline{f}(\hat{\underline{x}}(n|n), n)$$
$$\hat{\underline{x}}(n+2|n) = \underline{f}(\hat{\underline{x}}(n+1|n), n+1)$$
$$\vdots$$
$$\hat{\underline{x}}(n+m|n) = \underline{f}(\hat{\underline{x}}(n+m-1|n), n+m-1) \qquad (2.12)$$

Several investigators (see [13] and [29]) have pointed out that extrapolation of state variables is an interpretation of what the Wiener predictor does in the linear case.

For the more general system (2.3), the basic process is still Markovian. In this case a modified version of (2.9) is obtained

$$I_n = \tfrac{1}{2}\| \underline{x}(o)-\underline{m} \|^2_{\underline{P}^{-1}(o)} + \sum_{k=o}^{n} \tfrac{1}{2}\| \underline{z}(k)-\underline{h}(\underline{x}(k),k) \|^2_{\underline{R}^{-1}(k)}$$

$$+ \sum_{k=o}^{n-1} \left\{ \tfrac{1}{2}\| \underline{w}(k) \|^2_{\underline{Q}^{-1}(k)} + \underline{\lambda}'(k) \left[\underline{x}(k+1)-\underline{f}(\underline{x}(k),\underline{w}(k),k)\right] \right\} \qquad (2.13)$$

For this case (2.12) becomes

$$\hat{\underline{x}}(n+1|n) = \underline{f}(\hat{\underline{x}}(n|n), \underline{0}, n) \qquad (2.14)$$

2.5 Dynamic Programming Formulation

Having reduced the estimation problem to a minimization problem, the possibility arises of proceeding sequentially by a dynamic programming formulation [3, 4, 5, 20] and at each step obtaining an estimate of the present state. This can be done is either $\underline{f}^{-1}$ exists or $\underline{GQG}'$ is non-singular.

Case i. $\underline{f}^{-1}$ exists for $k=o,\ldots,n-1$

If $\underline{f}^{-1}$ exists we have from (2.1)

$$\underline{x}(k-1) = \underline{f}^{-1}(\underline{x}(k)-\underline{G}(k-1)\underline{w}(k-1),k-1) \tag{2.15}$$

In order to establish a sequential procedure, we define the following scalar "cost" function:

$$V_o(\underline{x}(o)) = \|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)} + \|\underline{z}(o)-\underline{h}(\underline{x}(o),o)\|^2_{\underline{R}^{-1}(o)}$$

$$V_n(\underline{x}(n)) = \underset{\underline{w}(o),\ldots,\underline{w}(n-1)}{\text{Min}} \Bigg\{ \|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)}$$

$$+ \sum_{k=o}^{n} \|\underline{z}(k)-\underline{h}(\underline{x}(k),k)\|^2_{\underline{R}^{-1}(k)} + \sum_{k=o}^{n-1} \|\underline{w}(k)\|^2_{\underline{Q}^{-1}(k)} \Bigg\} \quad n=1,2,\ldots \tag{2.16}$$

subject to the constraint (2.15). The separability property of V_n enables us to rewrite this relation in the following form

$$V_n(\underline{x}(n)) = \underset{\underline{w}(n-1)}{\text{Min}} \Bigg\{ \underset{\underline{w}(o),\ldots,\underline{w}(n-2)}{\text{Min}} \Bigg[\|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)}$$

$$+ \sum_{k=o}^{n-1} \|\underline{z}(k)-\underline{h}(\underline{x}(k),k)\|^2_{\underline{R}^{-1}(k)} + \sum_{k=o}^{n-2} \|\underline{w}(k)\|^2_{\underline{Q}^{-1}(k)} \Bigg]$$

$$+ \|\underline{z}(n)-\underline{h}(\underline{x}(n),n)\|^2_{\underline{R}^{-1}(n)} + \|\underline{w}(n-1)\|^2_{\underline{Q}^{-1}(n-1)} \Bigg\} \quad n=1,2,\ldots$$

subject to the constraint (2.15). This is equivalent to

$$V_n(\underline{x}(n)) = \underset{\underline{w}(n-1)}{\text{Min}} \Bigg\{ V_{n-1}(\underline{x}(n-1)) + \|\underline{z}(n)-\underline{h}(\underline{x}(n),n)\|^2_{\underline{R}^{-1}(n)}$$

$$+ \|\underline{w}(n-1)\|^2_{\underline{Q}^{-1}(n-1)} \Bigg\}$$

again subject to (2.15). Finally, using (2.15), we obtain the functional equation

$$V_n(\underline{x}(n)) = \underset{\underline{w}(n-1)}{\text{Min}} \left\{ V_{n-1}(\underline{f}^{-1}(\underline{x}(n)-\underline{G}(n-1)\underline{w}(n-1),n-1)) \right.$$

$$\left. + \| \underline{w}(n-1) \|^2_{\underline{Q}^{-1}(n-1)} + \| \underline{z}(n)-\underline{h}(\underline{x}(n),n) \|^2_{\underline{R}^{-1}(n)} \right\} \qquad (2.17)$$

Let $\underline{c}$ be any particular value of $\underline{x}(n)$. We may interpret $V_n(\underline{c})$ as a measure of the unlikeliness of the most probable sequence of states $\{\underline{x}(o),\ldots,\underline{x}(n)\}$ in which $\underline{x}(n)$ takes on the particular value $\underline{c}$, given the observed sequence $\{\underline{z}(o),\ldots,\underline{z}(n)\}$ and the a priori distribution for $\underline{x}(o)$. The estimate of $\underline{x}(n)$ is that value of $\underline{x}(n)$ for which $V_n(\underline{x}(n))$ is minimum. This corresponds to choosing the $\underline{x}(n)$ of the most probable (least unlikely) of all possible sequences of states given the observation and the a priori distribution.

This procedure may be extended to the more general system (2.3) if $\underline{x}(n-1)$ is uniquely specified by $\underline{x}(n)$ and $\underline{w}(n-1)$. For such systems there will exist a relation of the form

$$\underline{x}(n-1) = \underline{g}(\underline{x}(n),\underline{w}(n-1),n-1) \qquad (2.18)$$

Using (2.18) in place of (2.15) in the preceding development, the following functional equation is obtained;

$$V_n(\underline{x}(n)) = \underset{\underline{w}(n-1)}{\text{Min}} \left\{ V_{n-1}(\underline{g}(\underline{x}(n),\underline{w}(n-1),n-1)) \right.$$

$$\left. + \| \underline{w}(n-1) \|^2_{\underline{Q}^{-1}(n-1)} + \| \underline{z}(n)-\underline{h}(\underline{x}(n),n) \|^2_{\underline{R}^{-1}(n)} \right\} \qquad (2.19)$$

Case ii. $\underline{G}(k)\underline{Q}(k)\underline{G}'(k)$ is non-singular for $k=o,\ldots,n$.

Let us first consider the problem of predicting one step ahead. Then the function to be minimized is

$$2J_{n,1} = \sum_{k=o}^{n} \left\{ \| \underline{z}(k)-\underline{h}(\underline{x}(k),k) \|^2_{\underline{R}^{-1}(k)} + \| \underline{x}(k+1)-\underline{f}(\underline{x}(k),k) \|^2_{[\underline{GQG}']^{-1}} \right\}$$

$$+ \| \underline{x}(o)-\underline{m} \|^2_{\underline{P}^{-1}(o)}$$

We define a "cost" function as follows;

$$W_o(\underline{x}(o)) = \| \underline{x}(o)-\underline{m} \|^2_{\underline{P}^{-1}(o)}$$

$$W_{n+1}(\underline{x}(n+1)) = \underset{\underline{x}(o),\ldots,\underline{x}(n)}{\text{Min}} \left\{ \| \underline{x}(o)-\underline{m} \|^2_{\underline{P}^{-1}(o)} + \sum_{k=o}^{n} \left[\| \underline{z}(k)-\underline{h}(\underline{x}(k),k) \|^2_{\underline{R}^{-1}(k)} \right.\right.$$

$$\left.\left. + \| \underline{x}(k+1)-\underline{f}(\underline{x}(k),k) \|^2_{[\underline{GQG}']^{-1}} \right] \right\} \qquad n=1,2,\ldots \qquad (2.20)$$

Using the separability property as before, this relation may be re-written in the following form;

$$W_{n+1}(\underline{x}(n+1)) = \underset{\underline{x}(n)}{\text{Min}} \left\{ \underset{\underline{x}(o),\ldots,\underline{x}(n-1)}{\text{Min}} \left[\| \underline{x}(o)-\underline{m} \|^2_{\underline{P}^{-1}(o)} + \sum_{k=o}^{n-1} \left\{ \| \underline{z}(k)-\underline{h}(\underline{x}(k),k) \|^2_{\underline{R}^{-1}(k)} \right.\right.\right.$$

$$\left.\left. + \| \underline{x}(k+1)-\underline{f}(\underline{x}(k),k) \|^2_{[\underline{GQG}']^{-1}} \right\} \right] + \| \underline{z}(n)-\underline{h}(\underline{x}(n),n) \|^2_{\underline{R}^{-1}(n)}$$

$$\left. + \| \underline{x}(n+1)-\underline{f}(\underline{x}(n),n) \|^2_{[\underline{GQG}']^{-1}} \right\}$$

Using the definition of $W_n(\underline{x}(n))$, we obtain the following functional equation;

$$W_{n+1}(\underline{x}(n+1)) = \underset{\underline{x}(n)}{\text{Min}} \Big\{ W_n(\underline{x}(n)) + \| \underline{x}(n+1) - \underline{f}(\underline{x}(n),n) \|^2_{[\underline{GQG}']^{-1}}$$

$$+ \| \underline{z}(n) - \underline{h}(\underline{x}(n),n) \|^2_{\underline{R}^{-1}(n)} \Big\} \tag{2.21}$$

We may give $W_{n+1}(\underline{x}(n+1))$ an interpretation similar to the one given to $V_n(\underline{x}(n))$ in case i. Let $\underline{c}$ be any particular value of $\underline{x}(n+1)$. Then $W_{n+1}(\underline{c})$ is a measure of the unlikeliness of the most probable sequence of states $\{\underline{x}(o),\ldots,\underline{x}(n+1)\}$ in which $\underline{x}(n+1)$ takes on the particular value $\underline{c}$, given the observed sequence $\{\underline{z}(o),\ldots,\underline{z}(n)\}$ and the a priori distribution for $\underline{x}(o)$. The estimate of $\underline{x}(n+1)$ based on the observed sequence $\{\underline{z}(o),\ldots, \underline{z}(n)\}$ is that value of $\underline{x}(n+1)$ for which $W_{n+1}(\underline{x}(n+1))$ is minimum.

The estimate of $\underline{x}(n)$ based on $\{\underline{z}(o),\ldots,\underline{z}(n)\}$ is the value of $\underline{x}(n)$ for which the following function is minimum;

$$V_n(\underline{x}(n)) = W_n(\underline{x}(n)) + \| \underline{z}(n) - \underline{h}(n),n) \|^2_{\underline{R}^{-1}(n)} \tag{2.22}$$

where $V_n(\underline{x}(n))$ has the same interpretation as in case i. Since the value of $\underline{x}(n)$ for which $W_n(\underline{x}(n))$ is minimum is the predicted value of $\underline{x}(n)$ given $\{\underline{z}(o),\ldots,\underline{z}(n-1)\}$, we may give (2.22) the following simple intuitive interpretation. The new observation $\underline{z}(n)$ is weighted according to its reliability, $\underline{R}^{-1}$, and used to modify the predicted value of $\underline{x}(n)$ to obtain the current estimate of $\underline{x}(n)$. The degree of uncertainty of prior knowledge is reflected in the sensitivity of $V_n(\underline{x}(n))$ to changes in $\underline{x}(n)$.

If the basic system were linear, either of the two functional equations, (2.17) or (2.21), could be solved analytically by assuming a solution of the following form;

$$\| \underline{x}(n)-\hat{\underline{x}}(n) \|^2_{\underline{A}(n)} + c(n)$$

where $\underline{A}(n)$ is positive definite and is the inverse of the covariance matrix of the conditional distribution for $\underline{x}(n)$ given the observations, and where $\hat{\underline{x}}(n)$ is the estimate and $c(n)$ is a scalar. The recurrence relations obtained by Kalman [31] for the estimate and the covariance matrix can be obtained in this manner, but the algebra is quite involved. For a linear system, as we shall see shortly, an analytic solution may be obtained by other methods when the restrictions of cases i and ii above do not hold. In the literature [29, 31, 32, 25] it has been assumed that the transition matrix $\underline{F}(k)$ is non-singular which corresponds to case i.

It is perhaps worthwhile to review briefly how the computation would proceed using the functional equations. For greater detail the reader is referred to [3, 4, 5].

For (2,17) and (2.19) $\underline{w}$ is said to be the decision vector. This is because an appropriate "decision", $\underline{w}(n-1)$, may be associated with each possible state $\underline{x}(n)$ in such a manner that (2.17) or (2.19) is satisfied. For (2.21) the situation is slightly different in that $\underline{x}$ is both the decision vector and the state vector. From (2.21) we see that with each possible state $\underline{x}(n)$ we may associate an appropriate decision $\underline{x}(n-1)$. As the computation proceeds at each step the value of the "cost" function and the appropriate value of the decision vector may be calculated for each possible value of the state vector. An estimate is obtained by choosing the value of the state vector for which the "cost" function is minimum.

If one is interested only in estimating present state variables or predicting future state variables, there is no need to store old values of the decision vector for each possible value of the state vector. This is different from most dynamic programming problems.

If it is also of interest to improve the original estimates of past states after the entire observation sequence has been observed, then the sequence of appropriate decisions for each possible state must be stored. Trading time for memory, this may be done on tape. The revised estimate of $\underline{x}(k-1)$ is specified by the decision vector associated with the revised estimate of $\underline{x}(k)$.

The importance of the dynamic programming formulation of the estimation problem lies in the possibility of processing each new observation as it occurs. Herein there is hope of solving the estimation problem in real time. Modern computers have adequate fast memory to handle second order systems in this manner. For higher order systems approximation techniques, such as polynomial approximations, may be used in conjunction with the functional equations. Computational aspects of dynamic programming are discussed in [5].

2.6 Two-Point Boundary Value Problems

Let us now consider the problem of minimizing (2.9) which is rewritten here for convenience,

$$I_n = \tfrac{1}{2} \| \underline{x}(o)-\underline{m} \|^2_{\underline{P}^{-1}(o)} + \sum_{k=o}^{n} \tfrac{1}{2} \| \underline{z}(k)-\underline{h}(\underline{x}(k),k) \|^2_{\underline{R}^{-1}(k)}$$

$$+ \sum_{k=o}^{n-1} \left\{ \tfrac{1}{2} \| \underline{w}(k) \|^2_{\underline{Q}^{-1}(k)} + \underline{\lambda}'(k) \left[\underline{x}(k+1)-\underline{f}(\underline{x}(k),k)-\underline{G}(k)\underline{w}(k) \right] \right\} \tag{2.9}$$

Setting equal to zero the partial derivatives of I_n with respect to $\underline{w}(k)$, $\underline{\lambda}(k)$, (for $k=o,\ldots,n-1$), and $\underline{x}(k)$ (for $k=o,\ldots,n$) yields the following equations which $\hat{\underline{x}}(k|n)$ must satisfy;

$$\hat{\underline{x}}(k+1|n) = \underline{f}(\hat{\underline{x}}(k|n),k) + \underline{G}(k)\underline{w}(k) \qquad k=o,\ldots,n-1 \tag{2.23}$$

$$\underline{w}(k) = \underline{Q}(k)\underline{G}'(k)\underline{\lambda}(k) \qquad k=o,\ldots,n-1 \tag{2.24}$$

$$\underline{\lambda}(k-1) = (\partial \underline{f}'(\hat{\underline{x}}(k|n),n)/\partial \underline{x})\, \underline{\lambda}(k) + (\partial \underline{h}'(\hat{\underline{x}}(k|n),k)/\partial \underline{x})$$

$$\underline{R}^{-1}(k)\ [\,\underline{z}(k) - \underline{h}(\hat{\underline{x}}(k|n),n)\,] \quad k=1,\ldots,n \tag{2.25}$$

$$\underline{\lambda}(n) = \underline{o} \tag{2.26}$$

$$\hat{\underline{x}}(o|n) = \underline{m} + \underline{P}(o)\ (\partial \underline{h}'(\hat{\underline{x}}(o|n),o)/\partial \underline{x})\ \underline{R}^{-1}(o)\ [\,\underline{z}(o)-\underline{h}(\hat{\underline{x}}(o|n),o)\,]$$

$$+ \underline{P}(o)\ (\partial \underline{f}'(\hat{\underline{x}}(o|n),o)/\partial \underline{x})\ \underline{\lambda}(o) \tag{2.27}$$

Combining (2.23) and (2.24) to eliminate $\underline{w}$, we obtain necessary conditions in the form of the following two-point boundary value problem;

$$\hat{\underline{x}}(k+1|n) = \underline{f}(\hat{\underline{x}}(k|n),k) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k)\ \underline{\lambda}(k) \quad k=o,\ldots,n-1 \tag{2.28}$$

The equation for $\underline{\lambda}$ is (2.25) and the boundary conditions are given by (2.26) and (2.27). The more general problem of minimizing (2.13) yields the following similar two-point boundary value problem;

$$\hat{\underline{x}}(k+1|n) = \underline{f}(\hat{\underline{x}}(k|n),\underline{w}(k),k) \qquad k=o,\ldots,n-1 \tag{2.29}$$

$$\underline{w}(k) = \underline{Q}(k)\ (\partial \underline{f}'(\hat{\underline{x}}(k|n),\underline{w}(k))/\partial \underline{w})\ \underline{\lambda}(k) \qquad k=o,\ldots,n-1 \tag{2.30}$$

The remaining equations are identical to (2.25), (2.26) and (2.27) except that $(\partial \underline{f}(\hat{\underline{x}}(k|n),k)/\partial \underline{x})$ is replaced by $(\partial \underline{f}(\hat{\underline{x}}(k|n),\underline{w}(k),k)/\partial \underline{x})$ in (2.25) and (2.27).

These two-point boundary value problems are discrete analogues of the one obtained by Bryson and Frazier [10].

2.7 The Linear Problem

The linear version of the basic system is

$$\underline{x}(k+1) = \underline{F}(k)\underline{x}(k) + \underline{G}(k)\underline{w}(k) \tag{2.31}$$

$$\underline{z}(k) = \underline{H}(k)\underline{x}(k) + \underline{v}(k) \tag{2.32}$$

In this special case the two-point boundary value problem is

$$\underline{\hat{x}}(k+1|n) = \underline{F}(k)\underline{\hat{x}}(k|n) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k)\,\underline{\lambda}(k) \qquad k=o,\ldots,n-1 \tag{2.33}$$

$$\underline{\lambda}(k-1) = \underline{F}'(k)\,\underline{\lambda}(k) + \underline{H}'(k)\underline{R}^{-1}(k)\,[\,\underline{z}(k)-\underline{H}(k)\underline{\hat{x}}(k|n)] \quad k=o,\ldots,n-1 \tag{2.34}$$

with the boundary conditions

$$\underline{\hat{x}}(o|n) = \underline{m} + \underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o)\,[\,\underline{z}(o)-\underline{H}(o)\underline{\hat{x}}(o|n)] + \underline{P}(o)\underline{F}'(o)\,\underline{\lambda}(o) \tag{2.35}$$

$$\underline{\lambda}(n) = \underline{o} \tag{2.36}$$

2.71 Preliminary Discussion

Before presenting the general solution to this problem, we shall first consider the simple cases n=o and n=1 in order to develop some feeling for the nature of the solution.

For n=o, (2.35) and (2.36) may be combined to give

$$\underline{\hat{x}}(o|o) = \underline{m} + \underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o)\,[\,\underline{z}(o)-\underline{H}(o)\underline{\hat{x}}(o|o)] \tag{2.37}$$

or

$$[\underline{I} + \underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o)\underline{H}(o)]\ \hat{\underline{x}}(o|o) = \underline{m} + \underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o)\underline{z}(o)$$

The appropriate inverse will always exist for $\underline{R}(o)$ positive definite.

Thus,

$$\hat{\underline{x}}(o|o) = [\underline{I} + \underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o)\underline{H}(o)]^{-1}[\underline{m} + \underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o)\underline{z}(o)]$$

This may be rewritten in the more convenient form

$$\hat{\underline{x}}(o|o) = \underline{m} + [\underline{I} + \underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o)\underline{H}(o)]^{-1}\underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o)\ [\underline{z}(o)-\underline{H}(o)\underline{m}] \tag{2.38}$$

Let

$$\underline{C}(o) = [\underline{I} + \underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o)\underline{H}(o)]^{-1}\underline{P}(o) \tag{2.39}$$

Then (2.38) becomes

$$\hat{\underline{x}}(o|o) = \underline{m} + \underline{C}(o)\underline{H}'(o)\underline{R}^{-1}(o)\ [\underline{z}(o)-\underline{H}(o)\underline{m}] \tag{2.40}$$

A comparison of the boundary condition (2.35) and (2.37) shows that (2.35) may be reduced by a similar procedure to the following form;

$$\hat{\underline{x}}(o|n) = \hat{\underline{x}}(o|o) + \underline{C}(o)\underline{F}'(o)\underline{\lambda}(o) \tag{2.41}$$

The relation (2.41) is the key to the general solution, as we shall see shortly.

For the case n=1, (2.33), (2.34) and (2.41) become respectively,

$$\hat{\underline{x}}(1|1) = \underline{F}(o)\hat{\underline{x}}(o|1) + \underline{G}(o)\underline{Q}(o)\underline{G}'(o)\underline{\lambda}(o) \tag{2.42}$$

$$\underline{\lambda}(o) = \underline{H}'(1)\underline{R}^{-1}(1)\ [\underline{z}(1)- \underline{H}\ \hat{\underline{x}}(1|1)] \tag{2.43}$$

$$\hat{\underline{x}}(o|1) = \hat{\underline{x}}(o|o) + \underline{C}(o)\underline{F}'(o)\underline{\lambda}(o) \tag{2.44}$$

The relations (2.42) and (2.44) may be combined as

$$\hat{\underline{x}}(1|1) = \underline{F}(o)\hat{\underline{x}}(o|o) + [\underline{F}(o)\underline{C}(o)\underline{F}'(o) + \underline{G}(o)\underline{Q}(o)\underline{G}'(o)] \; \underline{\lambda}(o) \tag{2.45}$$

Let

$$\underline{P}(1) = \underline{F}(o)\underline{C}(o)\underline{F}'(o) + \underline{G}(o)\underline{Q}(o)\underline{G}'(o) \tag{2.46}$$

Using (2.46) and substituting for $\underline{\lambda}(o)$ from (2.43), we rewrite (2.45) as,

$$\hat{\underline{x}}(1|1) = \underline{F}(o)\hat{\underline{x}}(o|o) + \underline{P}(1)\underline{H}'(1)\underline{R}^{-1}(1) \; [\underline{z}(1) - \underline{H}\,\hat{\underline{x}}(1|1)] \tag{2.47}$$

Let us note that $\underline{F}(o)\hat{\underline{x}}(o|o) = \hat{\underline{x}}(1|o)$ by (2.12), and compare (2.47) with (2.37). We come to the immediate conclusion that a procedure similar to the one used for n=o will yield an equation similar to (2.40). Let

$$\underline{C}(1) = [\; \underline{I} + \underline{P}(1)\underline{H}'(1)\underline{R}^{-1}(1)\underline{H}(1) \;]^{-1}\underline{P}(1) \tag{2.48}$$

Then the equation corresponding to (2.40) is

$$\hat{\underline{x}}(1|1) = \hat{\underline{x}}(1|o) + \underline{C}(1)\underline{H}'(1)\underline{R}^{-1}(1) \; [\underline{z}(1) - \underline{H}(1)\hat{\underline{x}}(1|o)] \tag{2.49}$$

Equations (2.41), (2.46), (2.48) and (2.49) will now be shown to be similar in form to the solution for arbitrary n.

2.72. Solution to the Linear Two-Point Boundary Value Problem

Theorem: Consider the linear two-point boundary value problem given by (2.33), (2.34), (2.35) and (2.36) where $\underline{P}(o)$ and $\underline{R}(k)$ are positive definite for k=o,...,n. Then $\hat{\underline{x}}(k|n)$ satisfies the relation

$$\hat{\underline{x}}(k|n) = \hat{\underline{x}}(k|k) + \underline{C}(k)\underline{F}'(k)\underline{\lambda}(k) \qquad k=o,\dots,n \tag{2.50}$$

where $\hat{\underline{x}}(o|o)$ is given by (2.40) and

$$\underline{C}(k) = [\underline{I} + \underline{P}(k)\underline{H}'(k)\underline{R}^{-1}(k)\underline{H}(k)]^{-1}\underline{P}(k) \tag{2.51}$$

$$\underline{P}(k+1) = \underline{F}(k)\underline{C}(k)\underline{F}'(k) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k) \tag{2.52}$$

$$\hat{\underline{x}}(k|k) = \underline{F}(k-1)\hat{\underline{x}}(k-1|k-1) + \underline{C}(k)\underline{H}'(k)\underline{R}^{-1}(k)$$

$$[\underline{z}(k)-\underline{H}(k)\underline{F}(k-1)\hat{\underline{x}}(k-1|k-1)] \qquad k=1,\ldots,n \tag{2.53}$$

To establish a proof we proceed inductively showing that if (2.50) is satisfied by $\hat{\underline{x}}(k|n)$, then it is also satisfied for $\hat{\underline{x}}(k+1|n)$.

Proof: Suppose that $\hat{\underline{x}}(k|n)$ satisfies (2.50) for some $k \in (o,\ldots,n-1)$. Then, by (2.33),

$$\hat{\underline{x}}(k+1|n) = \underline{F}(k)\hat{\underline{x}}(k|k) + [\underline{F}(k)\underline{C}(k)\underline{F}'(k) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k)]\ \underline{\lambda}(k)$$

Substituting $\underline{P}(k+1)$ from (2.52) and using (2.34), we obtain

$$\hat{\underline{x}}(k+1|n) = \underline{F}(k)\hat{\underline{x}}(k|k) + \underline{P}(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)\ [\underline{z}(k+1)-\underline{H}(k+1)\hat{\underline{x}}(k+1|n)]$$

$$+ \underline{P}(k+1)\underline{F}'(k+1)\underline{\lambda}(k+1)$$

or

$$[\underline{I} + \underline{P}(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)\underline{H}(k+1)]\ \hat{\underline{x}}(k+1|n) = \underline{F}(k)\hat{\underline{x}}(k|k)$$

$$+ \underline{P}(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)\underline{z}(k+1) + \underline{P}(k+1)\underline{F}'(k+1)\underline{\lambda}(k+1)$$

The needed inverse will always exist. After some manipulation, this yields the following equation;

$$\hat{\underline{x}}(k+1|n) = \underline{F}(k)\hat{\underline{x}}(k|k) + \underline{C}(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)\ [\underline{z}(k+1)$$

$$- \underline{H}(k+1)\underline{F}(k)\hat{\underline{x}}(k|k)] + \underline{C}(k+1)\underline{F}'(k+1)\underline{\lambda}(k+1)$$

Using (2.53), we obtain the desired result;

$$\hat{\underline{x}}(k+1|n) = \hat{\underline{x}}(k+1|k+1) + \underline{C}(k+1)\underline{F}'(k+1)\underline{\lambda}(k+1)$$

To complete the proof we note that, by (2.41), the hypothesis is satisfied for $\hat{\underline{x}}(o|n)$.

Discussion: Note that no assumption is made that $\underline{F}(k)$ is non-singular. The recurrence relations (2.51), (2.52) and (2.53) were first obtained by Kalman [31] in a somewhat different form and using a very different approach. Using the following matrix identities derived in Appendix A;

$$[\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H}]^{-1}\, \underline{PH}'\underline{R}^{-1} = \underline{PH}'\,[\underline{HPH}' + \underline{R}]^{-1}$$

$$[\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H}]^{-1}\, \underline{P} = \underline{P} - \underline{PH}'\,[\underline{HPH}' + \underline{R}]^{-1}\,\underline{HP}$$

we may combine (2.51) and (2.52) and rewrite (2.53) to obtain equations more closely resembling Kalman's;

$$\hat{\underline{x}}(k|k) = \hat{\underline{x}}(k|k-1) + \underline{P}(k)\underline{H}'(k)\,[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)]^{-1}\,[\underline{z}(k)-\underline{H}(k)\hat{\underline{x}}(k|k-1)] \qquad (2.54)$$

$$\underline{P}(k+1) = \underline{F}(k)\,\{\underline{P}(k)-\underline{P}(k)\underline{H}'(k)\,[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)]^{-1}\,\underline{H}(k)\underline{P}(k)\}\,\underline{F}'(k) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k) \qquad (2.55)$$

For the linear system, (2.31) and (2.32), all probability distributions remain Gaussian. $\underline{P}(k)$ is the covariance matrix of the distribution for $\underline{x}(k)$ given $\{\underline{z}(o),\ldots,\underline{z}(k-1)\}$ and $\underline{C}(k)$ is the covariance matrix of the distribution for $\underline{x}(k)$ given $\{\underline{z}(o),\ldots,\underline{z}(k)\}$. (see Appendix D)

The solution of the problem of estimating the present state of the system is given by (2.51), (2.52) and (2.53), or equivalently by (2.51), (2.54) and (2.55). In practice the decision to use either (2.52) and

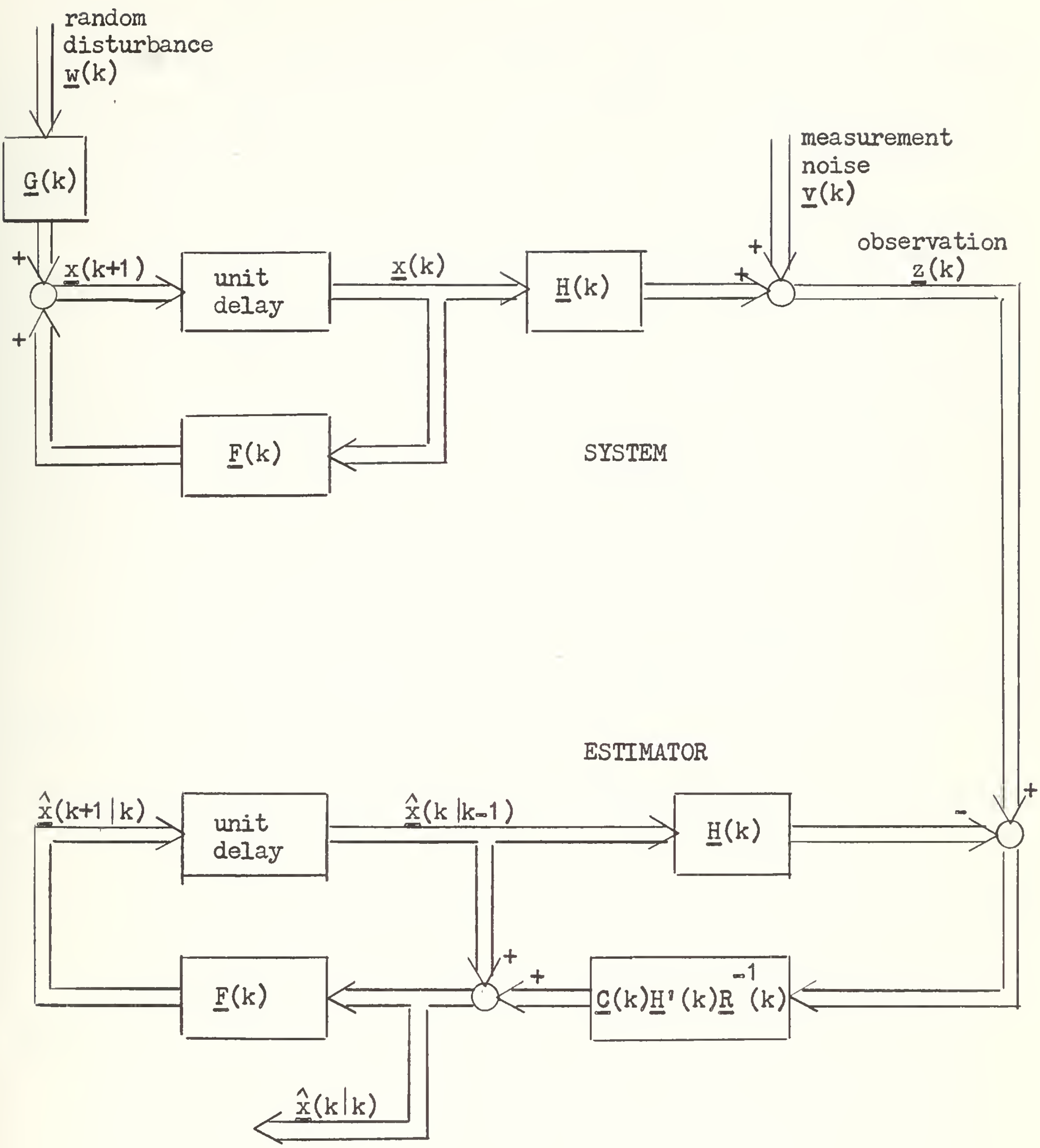


Fig. 7 Noisy linear system and optimal estimator

(2.53) or the equivalent relations (2.54) and (2.55) would be made to minimize the dimension of the matrix to be inverted. A block diagram of the system and the estimator is given in Fig. 7. The optimal estimator is linear and consists of a model of the system in a feedback loop. Compare Fig. 7 with Fig. 4.

If the estimate, $\hat{\underline{x}}(k|n)$ is desired, one could proceed as follows;

1. Compute and save $\hat{\underline{x}}(k|k)$ and $\underline{C}(k)$ for $k=o,\ldots,n$ using (2.51), (2.52) and (2.53) or, equivalently (2.51), (2.54) and (2.55).
2. Calculate $\hat{\underline{x}}(k|n)$ by backward recursion of (2.34) and (2.50). This procedure begins by calculating $\underline{\lambda}(n-1)$ using (2.34). Then obtain $\hat{\underline{x}}(n-1|n)$ using (2.50). Use (2.34) to obtain $\underline{\lambda}(n-2)$ and then (2.50) to obtain $\hat{\underline{x}}(n-2|n)$. Continue this alternating procedure until $\hat{\underline{x}}(o|n)$ is obtained.

An alternate procedure would be to combine (2.50) and (2.34) to obtain the following equation for $\underline{\lambda}$;

$$\underline{\lambda}(k-1) = \left[\underline{I} - \underline{H}'(k)\underline{R}^{-1}(k)\underline{H}(k)\underline{C}(k) \right] \underline{F}'(k)\underline{\lambda}(k) + \underline{H}'(k)\underline{R}^{-1}(k) \left[\underline{z}(k)-\underline{H}(k)\hat{\underline{x}}(k|k)\right] \tag{2.56}$$

The following matrix identities can easily be derived by the methods of Appendix A;

$$\underline{I} - \underline{H}'\underline{R}^{-1}\underline{H} \left[\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H} \right]^{-1}\underline{P} = \left[\underline{I} + \underline{H}'\underline{R}^{-1}\underline{HP} \right]^{-1}$$

$$= \underline{I} - \underline{H}' \left[\underline{R} + \underline{HPH}' \right]^{-1}\underline{HP}$$

Using these identities, (2.56) may be rewritten in different forms. Then step 2 above can be replaced by the following;

2. Calculate $\underline{\lambda}(k)$ by backward recursion of a convenient form of (2.56). Then calculate $\hat{\underline{x}}(k|n)$ using (2.50).

By repeated use of matrix identities (2.56) may be written in the following form which later will be compared with other results;

$$\underline{\lambda}(k-1) = \left\{ \underline{I} - \underline{H}'(k) \left[\underline{R}(k) + \underline{H}(k)\underline{P}(k)\underline{H}'(k) \right]^{-1} \underline{H}(k)\underline{P}(k) \right\} \underline{F}'(k)\underline{\lambda}(k)$$

$$+ \underline{H}'(k) \left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k) \right]^{-1} \left[\underline{z}(k) - \underline{H}(k)\hat{\underline{x}}(k|k-1) \right] \quad (2.57)$$

The solution of the linear smoothing problem may be given in an even simpler form if we write the equations in terms of $\hat{\underline{x}}(k|k-1)$. Later when we discuss the application of linearization techniques to nonlinear estimation problems, the formulas we shall obtain will be analogous to the ones already given for the linear problem, since we shall wish to linearize about the point $\hat{\underline{x}}(k|k)$ rather than the point $\hat{\underline{x}}(k|k-1)$.

Combining (2.50) and (2.54) using the matrix identities, we obtain the following equation;

$$\hat{\underline{x}}(k|n) = \hat{\underline{x}}(k|k-1) + \underline{P}(k)\underline{H}'(k) \left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k) \right]^{-1}$$

$$\left[\underline{z}(k) - \underline{H}(k)\hat{\underline{x}}(k|k-1) \right] + \underline{P}(k) \left\{ \underline{I} - \underline{H}'(k) \left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k) \right]^{-1} \right.$$

$$\left. \underline{H}(k)\underline{P}(k) \right\} \underline{F}'(k)\, \underline{\lambda}(k)$$

Combining this relation with (2.57) yields

$$\hat{\underline{x}}(k|n) = \hat{\underline{x}}(k|k-1) + \underline{P}(k)\underline{\lambda}(k-1)$$

We may summarize this result in the following corollary;

<u>Corollary</u>: The solution of the linear smoothing problem is given by the following equations;

$$\hat{\underline{x}}(k|n) = \hat{\underline{x}}(k|k-1) + \underline{P}(k)\underline{\lambda}(k-1)$$

$$\hat{\underline{x}}(k+1|k) = \underline{F}(k)\hat{\underline{x}}(k|k-1) + \underline{F}(k)\underline{P}(k)\underline{H}'(k)[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)]^{-1}$$

$$[\underline{z}(k) - \underline{H}(k)\hat{\underline{x}}(k|k-1)]$$

where $\underline{P}(k)$ is given by (2.55) and $\underline{\lambda}(k-1)$ is given by (2.57) for $k=o,\ldots,n$. The a priori mean $\underline{m}$ is used for $\hat{\underline{x}}(o|-1)$ in the above expressions and $\underline{\lambda}(n) = \underline{o}$.

2.73 The Existence of $[\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H}]^{-1}$

In the following discussion we shall use some matrix identities and the following elementary facts from matrix theory developed in Appendix A;

i. The inverse of a positive definite matrix is positive definite.

ii. The matrix formed by adding a positive definite matrix to a non-negative definite matrix is positive definite and hence, has an inverse.

One may verify by direct multiplication that the unique inverse of $\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H}$ is

$$\underline{I} - \underline{PH}' [\underline{HPH}' + \underline{R}]^{-1} \underline{H} \tag{2.58}$$

If $\underline{R}$ is positive definite and $\underline{P}$ is non-negative definite, (2.58) will always exist by virtue of ii. Using (2.58), $\underline{C}(k)$ may be written in the following form;

$$\underline{C}(k) = \underline{P}(k) - \underline{P}(k)\underline{H}'(k)[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)]^{-1}\underline{H}(k)\underline{P}(k) \tag{2.59}$$

The relation (2.59) is a generalization of a lemma given by Ho [25], who

required that $\underline{P}(k)$ be positive definite.

While the inverse of a matrix is unique, it frequently may be written in various forms which bear little resemblance to each other. If $\underline{P}(k)$ is positive definite for example, the inverse of $[\,\underline{I} + \underline{P}\underline{H}'\underline{R}^{-1}\underline{H}\,]$ may be written in the following form;

$$[\,\underline{P}^{-1} + \underline{H}'\underline{R}^{-1}\underline{H}\,]^{-1}\underline{P}^{-1} \tag{2.60}$$

which surprisingly is identical to (2.58). Using (2.60), we may write (2.51) in the following form;

$$\underline{C}(k) = [\,\underline{P}^{-1}(k) + \underline{H}'(k)\underline{R}^{-1}(k)\underline{H}(k)\,]^{-1} \tag{2.61}$$

From (2.61) and ii it follows that if $\underline{P}(k)$ is positive definite, then $\underline{C}(k)$ is positive definite. This is in accord with the fact that $\underline{P}(k)$ and $\underline{C}(k)$ are the covariance matrices of the probability distributions for $\underline{x}(k)$ given $\{\underline{z}(o),\ldots,\underline{z}(k-1)\}$ and $\{\underline{z}(o),\ldots,\underline{z}(k)\}$ respectively. A singular covariance matrix signifies that $\underline{x}(k)$ is confined with probability one to a hyperplane in n-space. It is clear that this cannot happen as the result of measurements which are corrupted with noise having a non-singular covariance matrix $\underline{R}(k)$.

Consider (2.52), which we rewrite for convenience,

$$\underline{P}(k+1) = \underline{F}(k)\underline{C}(k)\underline{F}'(k) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k) \tag{2.52}$$

From (2.52) we see that a sufficient condition for $\underline{P}(k+1)$ to be positive definite is that either $\underline{G}(k)\underline{Q}(k)\underline{G}'(k)$ be positive definite or $\underline{C}(k)$ be positive definite and $\underline{F}(k)$ be non-singular. This may easily be related to the basic equation for the system,

$$\underline{x}(k+1) = \underline{F}(k)\underline{x}(k) + \underline{G}(k)\underline{w}(k) \tag{2.31}$$

From (2.31) we see that $\underline{P}(k+1)$ will be positive definite if every point in n-space may be reached in one transition from some point in the set of possible values of $\underline{x}(k)$ (specified by $\underline{C}(k)$) using some $\underline{w}(k)$. Thus, the condition obtained from (2.52) is seen to be sufficient but not necessary.

Finally, we note that if $\underline{P}(o)$, $\underline{R}(k)$ and $\underline{F}(k)$ are non-singular for all k, then $\underline{P}(k)$ and $\underline{C}(k)$ are also positive definite for all k.

2.74 Generalizations

The restrictions that $\underline{P}(o)$ and $\underline{R}(k)$ be positive definite may be relaxed without altering the basic structure of the solution. In order to relax these restrictions we use the generalized inverse discussed in Appendix B. Appropriate background material concerning probabilistic aspects of singular covariance matrices is given in Appendix C. In this section we shall simply state the results obtained in Appendix D and compare them with those already obtained in this chapter. For a more detailed discussion the reader is referred to Appendix D, which considers a very general version of the linear estimation problem in which correlation between $\underline{v}$ and $\underline{w}$ is allowed and the covariance matrices $\underline{P}(o)$ and $\underline{R}(k)$ may be singular.

The solution of the linear estimation problem is given by the following equations which are the same as (2.50), (2.59), (2.52), (2.54) and (2.57) except that in (2.59), (2.54) and (2.57) the inverse is replaced by the generalized inverse;

$$\hat{\underline{x}}(k|n) = \hat{\underline{x}}(k|k) + \underline{C}(k)\underline{F}'(k)\underline{\lambda}(k) \qquad k=o,\ldots,n \tag{2.50}$$

$$\underline{C}(k) = \underline{P}(k)-\underline{P}(k)\underline{H}'(k)\ [\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)]^{\#}\underline{H}(k)\underline{P}(k) \tag{2.62}$$

$$\underline{P}(k+1) = \underline{F}(k)\underline{C}(k)\underline{F}'(k) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k) \tag{2.52}$$

$$\underline{\hat{x}}(k|k) = \underline{\hat{x}}(k|k-1) + \underline{P}(k)\underline{H}'(k)\left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)\right]^{\#}$$

$$\left[\underline{z}(k)-\underline{H}(k)\underline{\hat{x}}(k|k-1)\right] \qquad k=1,\ldots,n \qquad (2.63)$$

$$\underline{\lambda}(k-1) = \left\{\underline{I} - \underline{H}'(k)\left[\underline{R}(k) + \underline{H}(k)\underline{P}(k)\underline{H}'(k)\right]^{\#}\underline{H}(k)\underline{P}(k)\right\}\underline{F}'(k)\underline{\lambda}(k)$$

$$+ \underline{H}'(k)\left[\underline{R}(k) + \underline{H}(k)\underline{P}(k)\underline{H}'(k)\right]^{\#}\left[\underline{z}(k)-\underline{H}(k)\underline{\hat{x}}(k|k-1)\right]$$

$$k=1,\ldots,n \qquad (2.64)$$

The boundary conditions are

$$\underline{x}(o|o) = \underline{m} + \underline{P}(o)\underline{H}'(o)\left[\underline{H}(o)\underline{P}(o)\underline{H}'(o) + \underline{R}(o)\right]^{\#}\left[\underline{z}(o) - \underline{H}(o)\underline{m}\right] \quad (2.65)$$

$$\underline{\lambda}(n) = \underline{o}$$

The generalized inverse may be replaced by any pseudo-inverse. The pseudo-inverse was used by Kalman [29] in his solution of the linear filtering and prediction problem. We have extended this work to include the smoothing problem and have relaxed the restriction that $\underline{F}(k)$ be non-singular. The only restrictions remaining are that $\underline{R}(k)$ and $\underline{P}(o)$ be valid covariance matrices and that observations do not occur which are inconsistent with prior knowledge. This point is discussed in Appendix C. The block diagram of Fig. 7, in which the estimator producing up-to-date estimates of present state variables consists of a model of the system in a linear feedback loop, remains valid if $\underline{C}(k)\underline{H}'(k)\underline{R}^{-1}(k)$ is replaced by

$$\underline{P}(k)\underline{H}'(k)\left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)\right]^{\#}$$

These two expressions are equivalent if $\underline{R}^{-1}$ exists.

If one is interested in the linear smoothing problem, the corollary of section 2.72 remains valid if the inverse in the equation for $\underline{\hat{x}}(k+1|k)$ is replaced by the generalized inverse.

2.8 An Approximation Technique

2.81 Linear Observations

There is a great need for approximation techniques which can be used to obtain estimates in real time. In this section we shall present a method which is computationally attractive and which possesses some intuitive appeal.

For simplicity we shall first assume that $\underline{h}(\underline{x})$ is linear, an assumption which will appreciably simplify the resulting equations and which represents a case of great practical significance. In this case we say that the observations are linear. Consider the approximation to (2.1) obtained by the following linearization;

$$\underline{x}(k+1) \approx \underline{f}(\underline{x}^*(k|k),k) + (\partial \underline{f}(\underline{x}^*(k|k),k)/\partial \underline{x}) \left[\underline{x}(k)-\underline{x}^*(k|k)\right] + \underline{G}(k)\underline{w}(k) \quad (2.66)$$

The point $\underline{x}^*(k|k)$ about which the linearization is made is the estimate of $\underline{x}(k)$ given $\{\underline{z}(o),\ldots,\underline{z}(k)\}$. We use the asterisk to remind ourselves that the estimates produced by this linearization procedure are only approximations and do not necessarily correspond exactly to the mode of the a posteriori distribution.

If the function I_n of (2.9) is replaced by

$$\begin{aligned} I_n^* = {} & \tfrac{1}{2}\left\|\underline{x}(o)-\underline{m}\right\|^2_{\underline{P}^{-1}(o)} + \sum_{k=o}^{n} \tfrac{1}{2}\left\|\underline{z}(k)-\underline{H}(k)\underline{x}(k)\right\|^2_{\underline{R}^{-1}(k)} \\ & + \sum_{k=o}^{n-1}\Big\{ \tfrac{1}{2}\left\|\underline{w}(k)\right\|^2_{\underline{Q}^{-1}(k)} + \underline{\lambda}'(k)\Big[\left[\underline{x}(k+1)-\underline{f}(\underline{x}^*(k|k),k)\right] \\ & - (\partial \underline{f}(\underline{x}^*(k|k),k)/\partial \underline{x}) \left[\underline{x}(k)-\underline{x}^*(k|k)\right] - \underline{G}(k)\underline{w}(k)\Big]\Big\} \end{aligned} \quad (2.67)$$

and the partial derivatives of I_n^*, with respect to $\underline{w}(k)$, $\underline{\lambda}(k)$ and $\underline{x}(k)$ are set equal to zero, a modified version of the two-point boundary value problem is obtained. The resulting

equations are as follows;

$$\underline{x}^*(k+1|n) = \underline{f}(\underline{x}^*(k|k),k) + (\partial \underline{f}(\underline{x}^*(k|k),k)/\partial \underline{x}) \left[\underline{x}^*(k|n)-\underline{x}^*(k|k)\right]$$

$$+ \underline{G}(k)\underline{Q}(k)\underline{G}'(k)\underline{\lambda}(k) \tag{2.68}$$

$$\underline{\lambda}(k-1) = (\partial \underline{f}'(\underline{x}^*(k|k),k)/\partial \underline{x}) \underline{\lambda}(k) + \underline{H}'(k)\underline{R}^{-1}(k) \left[\underline{z}(k)-\underline{H}(k)\underline{x}^*(k|n)\right] \tag{2.69}$$

The boundary conditions are

$$\underline{\lambda}(n) = \underline{o} \tag{2.70}$$

$$\underline{x}^*(o|n) = \underline{m} + \underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o) \left[\underline{z}(o)-\underline{H}(o)\underline{x}^*(o|n)\right]$$

$$+ \underline{P}(o)(\partial \underline{f}'(\underline{x}^*(o|o),o)/\partial \underline{x}) \underline{\lambda}(o) \tag{2.71}$$

A comparison of (2.71) and (2.35) indicates that (2.71) may be written in the following form which is similar to (2.41);

$$\underline{x}^*(o|n) = \underline{x}^*(o|o) + \underline{C}^*(o)(\partial \underline{f}'(\underline{x}^*(o|o),o)/\partial \underline{x}) \underline{\lambda}(o) \tag{2.72}$$

where $\underline{x}^*(o|o)$ and $\underline{C}^*(o)$ are equal to $\hat{\underline{x}}(o|o)$ and $\underline{C}(o)$ respectively and are given by (2.40) and (2.39). It is not at all surprising that the solution of this linearized problem closely resembles that of the linear problem already studied. We define the following quantities;

$$\underline{C}^*(k) = \left[\underline{I} + \underline{P}^*(k)\underline{H}'(k)\underline{R}^{-1}(k) \right]^{-1}\underline{P}^*(k) \tag{2.73}$$

$$\underline{P}^*(k+1) = (\partial \underline{f}(\underline{x}^*(k|k),k)/\partial \underline{x}) \underline{C}^*(k) (\partial \underline{f}'(\underline{x}^*(k|k),k)/\partial \underline{x})$$

$$+ \underline{G}(k)\underline{Q}(k)\underline{G}'(k) \tag{2.74}$$

$$\underline{P}^*(o) = \underline{P}(o) \tag{2.75}$$

We now hypothesize that the $\underline{x}^*(k|n)$ satisfying the two-point boundary value problem (2.68), (2.69), (2.70) and (2.71) is given by the following equations;

$$\underline{x}^*(k|n) = \underline{x}^*(k|k) + \underline{C}^*(k)(\partial \underline{f}'(\underline{x}^*(k|k),k)/\partial \underline{x})\ \underline{\lambda}(k) \qquad k=0,\ldots,n \quad (2.76)$$

where

$$\underline{x}^*(k|k) = \underline{f}(\underline{x}^*(k-1|k-1),k-1) + \underline{C}^*(k)\underline{H}'(k)\underline{R}^{-1}(k)$$

$$[\,\underline{z}(k)-\underline{H}(k)\underline{f}(\underline{x}^*(k-1|k-1),k-1)\,] \qquad k=1,\ldots,n \quad (2.77)$$

and $\underline{x}^*(o|o)$ is given by (2.40). We again proceed by induction showing that if (2.76) is satisfied for some $k \in (o,\ldots,n-1)$ then it is also satisfied for k+1. We know that (2.76) is satisfied for k=o. The steps in the following proof are completely analogous to those used to prove the theorem of section 2.72.

Suppose that $\underline{x}^*(k|n)$ satisfies (2.76) for some $k \in (o,\ldots,n-1)$. Then, by (2.68)

$$\underline{x}^*(k+1|n) = \underline{f}(\underline{x}^*(k|k),k) - (\partial \underline{f}(\underline{x}^*(k|k),k)/\partial \underline{x})\ \underline{x}^*(k|k)$$

$$+ \ (\partial \underline{f}(\underline{x}^*(k|k),k)/\partial \underline{x})\ \underline{x}^*(k|k)$$

$$+\Big\{ (\partial \underline{f}(\underline{x}^*(k|k),k)/\partial \underline{x})\ \underline{C}^*(k)\ (\partial \underline{f}'(\underline{x}^*(k|k),k)/\partial \underline{x})$$

$$+ \underline{G}(k)\underline{Q}(k)\underline{G}'(k)\Big\}\ \underline{\lambda}(k)$$

or, equivalently

$$\underline{x}^*(k+1|n) = \underline{f}(\underline{x}^*(k|k),k) + \underline{P}^*(k+1)\ \underline{\lambda}(k)$$

Substituting for $\underline{\lambda}(k)$ from (2.69) this expression becomes

$$\underline{x}^*(k+1|n) = \underline{f}(\underline{x}^*(k|k),k) + \underline{P}^*(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)$$

$$[\underline{z}(k+1) - \underline{H}(k)\underline{x}^*(k+1|n)]$$

$$+ \underline{P}^*(k+1)(\partial \underline{f}'(\underline{x}^*(k+1|k+1),k+1)/\partial \underline{x})\ \underline{\lambda}(k+1)$$

Solving this expression for $\underline{x}^*(k+1|n)$ yields

$$\underline{x}^*(k+1|n) = [\underline{I} + \underline{P}^*(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)\underline{H}(k+1)]^{-1}$$

$$\{\underline{f}(\underline{x}^*(k|k),k) + \underline{P}^*(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)\underline{z}(k+1)$$

$$+ \underline{P}^*(k+1)(\partial \underline{f}'(\underline{x}^*(k+1|k+1),k+1)/\partial \underline{x})\ \underline{\lambda}(k+1)\}$$

This expression may be rewritten in the following form;

$$\underline{x}^*(k+1|n) = \underline{f}(\underline{x}^*(k|k),k) + \underline{C}^*(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)$$

$$[\underline{z}(k+1)-\underline{H}(k+1)\underline{f}(\underline{x}^*(k|k),k)]$$

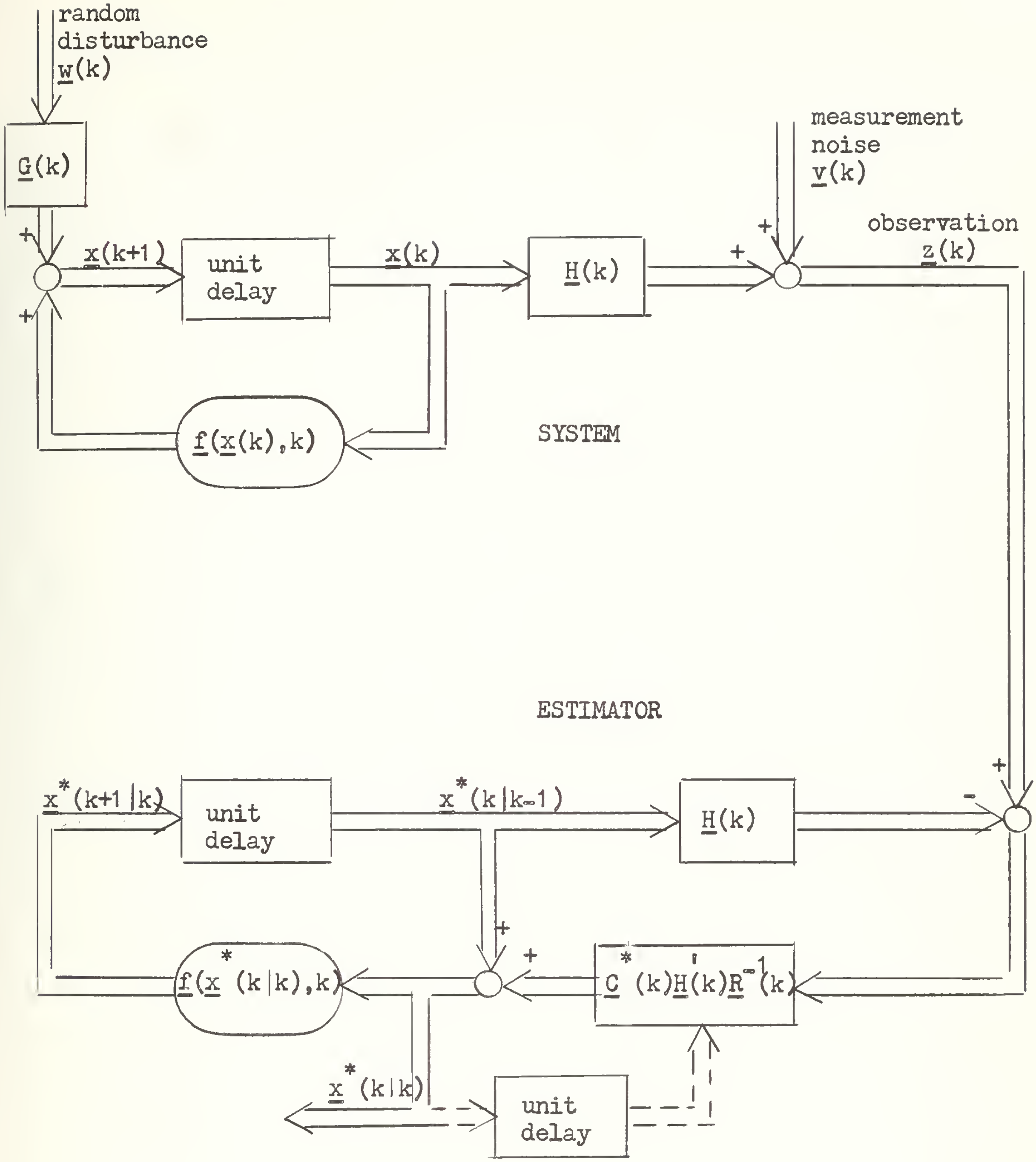
$$+ \underline{C}^*(k+1)(\partial \underline{f}'(\underline{x}^*(k+1|k+1),k+1)/\partial \underline{x})\ \underline{\lambda}(k+1)$$

Using (2.77), this expression becomes

$$\underline{x}^*(k+1|n) = \underline{x}^*(k+1|k+1) + \underline{C}^*(k+1)(\partial \underline{f}'(\underline{x}^*(k+1|k+1),k+1)/\partial \underline{x})\ \underline{\lambda}(k+1)$$

This expression satisfies the hypothesis (2.77) and noting that (2.77) is satisfied for k=o concludes the proof.

Equations (2.73), (2.74) and (2.77) constitute a recursive scheme in which each new observation can be processed immediately to produce an up-to-date estimate. The matrices $\underline{C}^*$ and $\underline{P}^*$ can no longer be strictly interpreted as covariance matrices of the distribution for $\underline{x}$, as they could be if the basic system were actually described by (2.66) instead of (2.1).



Dashed lines indicate $\underline{x}^*(k|k)$ is needed to calculate $\underline{C}^*(k+1)$

Fig. 8 Nonlinear discrete-time estimator

A block diagram of the estimator described by equation (2.77) is given in Fig. 8. Again we see the model of the system appearing in a feedback loop. Equation (2.77) or, equivalently, the estimator of Fig. 8 may be given a simple intuitive interpretation if we interpret the individual terms in the following manner;

$\underline{f}(\underline{x}^*(k-1\|k-1),k-1)$	a prediction of $\underline{x}(k)$
$\underline{C}^*(k)$	a measure of present uncertainty
$\underline{H}'(k)$	sensitivity of observation to deviations in $\underline{x}(k)$
$\underline{R}^{-1}(k)$	reliability of measurements
$\underline{H}(k)\underline{f}(\underline{x}^*(k-1\|k-1),k-1)$	a prediction of $\underline{z}(k)$

The equation then has the interpretation that the error in the prediction of $\underline{z}(k)$ is weighted in proportion to present uncertainty, sensitivity of the observation, and reliability of the measurement. This weighted error is then used to modify the predicted value of $\underline{x}(k)$ to obtain an up-to-date estimate of $\underline{x}(k)$.

Equations (2.73), (2.74) and (2.77) or equivalent equations obtained by the use of matrix identities are particularly suitable for real time implementation on a digital computer. Simulation studies may be conducted to evaluate the applicability of this approximation to a particular system. The results of some simulation studies using this linearized estimation technique are given in Chapter V.

When a detailed analysis of all available data is to be made after the completion of some experiment such as a rocket flight, one is faced with numerical solution of the nonlinear two-point boundary value problem. Some method of successive approximations, such as a version of the gradient method of Kelley [35] and Bryson [9], is usually called for. There is

danger of converging to a local minimum if the conditional probability density function (2.7) is not unimodal. This danger is reduced and the speed of convergence increased if a good first approximation is available. The logical choice for such a first approximation is $\underline{x}^*(k|n)$ which may be calculated by backward recursion of (2.69) and (2.76). Alternately, we may combine (2.69) and (2.76) to obtain the following equation for $\underline{\lambda}$ which is similar to (2.56);

$$\underline{\lambda}(k-1) = [\,\underline{I} - \underline{H}'(k)\underline{R}^{-1}(k)\underline{H}(k)\underline{C}^*(k)\,] \quad (\partial \underline{f}'(\underline{x}^*(k|k),k))/\partial \underline{x})\ \underline{\lambda}(k)$$

$$+ \underline{H}'(k)\underline{R}^{-1}(k)\ [\,\underline{z}(k) - \underline{H}(k)\underline{x}^*(k|k)\,] \tag{2.78}$$

Then $\underline{x}^*(k|n)$ may be calculated using (2.78) and (2.76).

2.82 Generalizations

Having obtained some confidence that linearization techniques will produce equations closely resembling the equations for the linear estimation problem, there are several directions in which one could generalize. First, one could complicate the problem without introducing new nonlinearities by allowing $\underline{R}(k)$ and $\underline{P}(o)$ to be singular and permitting correlation between $\underline{w}$ and $\underline{v}$. That is, one could study the problem of Appendix D using (2.66) in place of (2.1). Second, one could introduce new nonlinearities by allowing $\underline{h}$ to be nonlinear and/or allowing $\underline{w}$ to enter nonlinearly as in (2.3). Third, one could combine these first two types of complications.

We shall discuss each of these possibilities briefly, indicating how the results obtained previously would be modified in these situations, but omitting detailed derivations.

The generalization to include singular $\underline{R}(k)$ and $\underline{P}(o)$ and allow correlation between $\underline{v}$ and $\underline{w}$ is completely straightforward. The derivation

of these results is completely analogous to section D5 of Appendix D and may easily be obtained by using section D5 as a guide. The basic equations which result from such a derivation are as follows;

$$\underline{x}^*(k|k) = \underline{x}^*(k|k-1) + \underline{P}^*(k)\underline{H}'(k)\underline{B}^*(k)\ [\ \underline{z}(k)-\underline{H}(k)\underline{x}^*(k|k-1)] \tag{2.79}$$

$$\underline{x}^*(k+1|k) = \underline{f}(\underline{x}^*(k|k),k) + \underline{G}(k)\underline{S}(k)\underline{B}^*(k)\ [\ \underline{z}(k)-\underline{H}(k)\underline{x}^*(k|k-1)] \tag{2.80}$$

$$\underline{x}^*(k|n) = \underline{x}^*(k|k) + \underline{P}^*(k)\ [\ (\partial\, \underline{f}'(\underline{x}^*(k|k),k)/\partial\, \underline{x})-\underline{H}'(k)\underline{M}^*(k)\ ]\, \underline{\lambda}(k) \qquad k=o,\ldots,n \tag{2.81}$$

$$\begin{aligned}\underline{\lambda}(k-1) = {} & [\ (\partial\, \underline{f}'(\underline{x}^*(k|k),k)/\partial\, \underline{x}) - \underline{H}'(k)\underline{M}^*(k)]\ \underline{\lambda}(k) \\ & + \underline{H}'(k)\underline{B}^*(k)\ [\ \underline{z}(k)-\underline{H}(k)\underline{x}^*(k|k-1)] \qquad k=1,\ldots,n\end{aligned} \tag{2.82}$$

$$\begin{aligned}\underline{P}^*(k+1) = {} & (\partial\, \underline{f}(\underline{x}^*(k|k),k)/\partial \underline{x})\ \underline{P}^*(k)\ \ (\partial\, \underline{f}'(\underline{x}^*(k|k),k)/\partial\, \underline{x}) \\ & + \underline{G}(k)\underline{Q}(k)\underline{G}'(k) - \underline{M}^{*\prime}(k)\underline{B}^{*\#}(k)\underline{M}^*(k)\end{aligned} \tag{2.83}$$

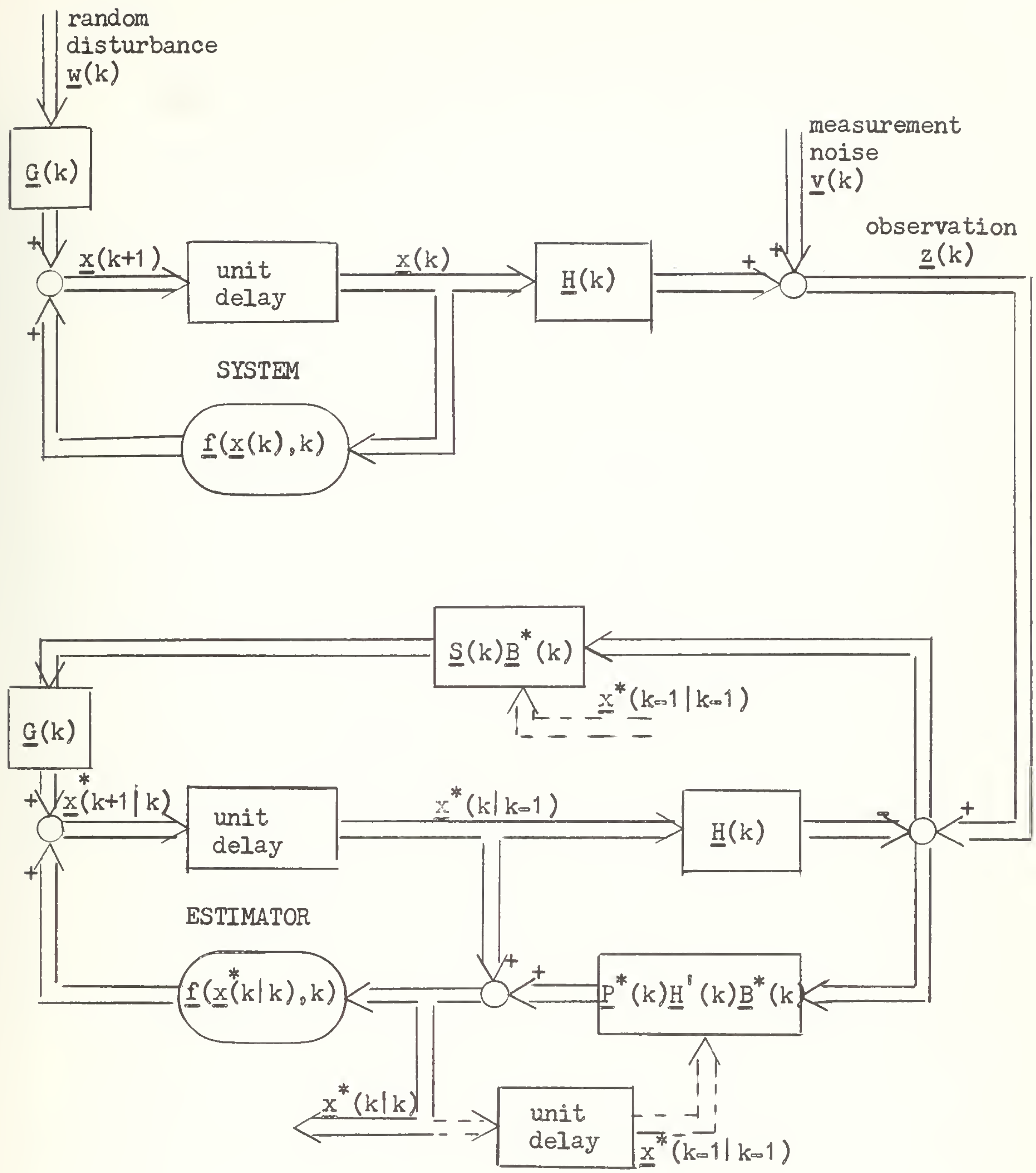
$$\underline{B}^*(k) = [\ \underline{H}(k)\underline{P}^*(k)\underline{H}'(k) + \underline{R}(k)]^{\#} \tag{2.84}$$

$$\underline{M}^*(k) = \underline{B}^*(k)\,[\,\underline{H}(k)\underline{P}^*(k)(\partial\, \underline{f}'(\underline{x}^*(k|k),k)/\partial\, \underline{x}) + \underline{S}'(k)\underline{G}'(k)] \tag{2.85}$$

A block diagram of the estimator described by equation (2.79) and (2.80) is given in Fig. 9. Again we have a model of the system appearing in a feedback loop.

If new nonlinearities are introduced, then we are forced to further linearization. We shall first consider the case in which $\underline{h}$ is nonlinear. In order to acquire a feeling for this situation we shall examine the simplest possible example, n=o. Then,

$$I_o = \tfrac{1}{2}\,\|\, \underline{x}(o)-\underline{m}\,\|^2_{\underline{P}^{-1}(o)} + \tfrac{1}{2}\,\|\, \underline{z}(o)-\underline{h}(\underline{x}(o),o)\,\|^2_{\underline{R}^{-1}(o)} \tag{2.86}$$



Dashed lines indicate $\underline{x}^*(k-1|k-1)$ is needed to calculate $\underline{P}^*(k)$ and $\underline{B}^*(k)$

Fig. 9 Nonlinear discrete-time estimator when $\underline{v}(k)$ and $\underline{w}(k)$ are correlated.

or, equivalently,

$$I_o = \tfrac{1}{2}\|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)} + \tfrac{1}{2}\|\underline{z}(o)-\underline{h}(\underline{m},o)-[\underline{h}(\underline{x}(o),o)-\underline{h}(\underline{m},o)]\|^2_{\underline{R}^{-1}(o)} \tag{2.87}$$

This may be rewritten as

$$I_o = \tfrac{1}{2}\|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)} + \tfrac{1}{2}\|\underline{z}(o)-\underline{h}(\underline{m},o)\|^2_{\underline{R}^{-1}(o)} + \tfrac{1}{2}\|\underline{h}(\underline{x}(o),o)-\underline{h}(\underline{m},o)\|^2_{\underline{R}^{-1}(o)}$$

$$-[\underline{h}(\underline{x}(o),o)-\underline{h}(\underline{m},o)]'\,\underline{R}^{-1}(o)\,[\underline{z}(o)-\underline{h}(\underline{m},o)] \tag{2.88}$$

The second term in this expression is independent of $\underline{x}$. The third and fourth terms we expand about the a priori mean $\underline{m}$ retaining the first non-zero terms. The fourth term becomes

$$(\underline{x}(o)-\underline{m})'(\partial\underline{h}'(\underline{m},o)/\partial\underline{x})\,\underline{R}^{-1}(o)\,[\underline{z}(o)-\underline{h}(\underline{m},o)] \tag{2.89}$$

The third term becomes

$$\tfrac{1}{2}(\underline{x}(o)-\underline{m})'\left\{\partial[(\partial\underline{h}'(\underline{m},o)/\partial\underline{x})\,\underline{R}^{-1}(o)\,\underline{h}(\underline{m},o)]/\partial\underline{x}\right\}(\underline{x}(o)-\underline{m}) \tag{2.90}$$

Let

$$\underline{C}^{-1}(o) = \underline{P}^{-1}(o) + \partial[(\partial\underline{h}'(m,o)/\partial\underline{x})\,\underline{R}^{-1}(o)\,\underline{h}(\underline{m},o)]/\partial\underline{x} \tag{2.91}$$

Then (2.86) may be approximated by the following expression

$$I_o \approx \tfrac{1}{2}\|\underline{x}(o)-\underline{m}\|^2_{\underline{C}^{-1}(o)} + \tfrac{1}{2}\|\underline{z}(o)-\underline{h}(\underline{m},o)\|^2_{\underline{R}^{-1}(o)}$$

$$+(\underline{x}(o)-\underline{m})'(\partial\underline{h}'(\underline{m},o)/\partial\underline{x})\,\underline{R}^{-1}(o)\,[\underline{z}(o)-\underline{h}(\underline{m},o)] \tag{2.92}$$

Differentiating this expression with respect to $\underline{x}(o)$ and setting the result equal to zero yields

$$\underline{x}^*(o|o) = \underline{m} + \underline{C}(o)(\partial\, \underline{h}'(\underline{m},o)/\partial\, \underline{x})\ \underline{R}^{-1}(o)\ \left[\underline{z}(o)-\underline{h}(\underline{m},o)\right] \tag{2.93}$$

Equations (2.91) and (2.93) are similar in structure to (2.72) and (2.77). If we interpret $(\partial\,\underline{h}'(\underline{m},o)/\partial\,\underline{x})$ as the sensitivity of the observations to deviations of $\underline{x}$ from the prior mean, we may give (2.93) and interpretation similar to that which was given in (2.77). The same basic structure is retained for arbitrary n. Note that $\underline{h}(\underline{x}(k),k)$ is linearized about $\underline{x}^*(k|k-1)$ since $\underline{x}^*(k|k)$ cannot be computed until $\underline{z}(k)$ is observed. The expression for $\underline{C}(k)$ is of the same form as the expression for $\underline{C}(o)$ given by (2.91). $\underline{P}(k+1)$ is computed by (2.74) as before.

Let us again consider the case n=o, but now allow $\underline{R}(o)$ to be singular. In this case, following Appendix D, we consider $\underline{v}$ to be the result of a linear transformation on a normalized random vector $\underline{u}$. That is, $\underline{v} = \underline{Tu}$ where $\underline{TT}'$ is equal to $\underline{R}(o)$. Introducing the Lagrange multiplier vector $\underline{\varrho}$, the expression to be minimized becomes

$$I_o = \tfrac{1}{2}\, \|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)} + \tfrac{1}{2}\,\|\underline{u}\|^2 + \underline{\varrho}'\ \left[\underline{z}(o)-\underline{h}(\underline{x}(o),o)-\underline{Tu}\right] \tag{2.94}$$

This expression will be approximated by the following linearized expression;

$$I_o^* = \tfrac{1}{2}\,\|\underline{x}(o)-\underline{m}\|^2_{\underline{P}^{-1}(o)} + \tfrac{1}{2}\,\|\underline{u}\|^2 + \underline{\varrho}'\left\{\underline{z}(o)-\underline{h}(\underline{m},o)-(\partial\,\underline{h}(\underline{m},o)/\partial\,\underline{x})\left[\underline{x}(o)-\underline{m}\right] - \underline{Tu}\right\} \tag{2.95}$$

Setting the total differential, $dI_o^*(\underline{x}(o),\underline{u},\underline{\varrho})$, equal to zero yields the

following equations;

$$\underline{x}^*(o|o) = \underline{m} + \underline{P}(o)(\partial \underline{h}'(\underline{m},o)/\partial \underline{x})\ \underline{\varrho} \tag{2.96}$$

$$\underline{z}(o) = \underline{h}(\underline{m},o) + (\partial \underline{h}(\underline{m},o)/\partial \underline{x})\ [\ \underline{x}^*(o|o)-\underline{m}\ ] + \underline{Tu} \tag{2.97}$$

$$\underline{u} = \underline{T}'\underline{\varrho} \tag{2.98}$$

Combining these three equations, we obtain

$$\underline{z}(o)-\underline{h}(\underline{m},o) = [\ (\partial \underline{h}(\underline{m},o)/\partial \underline{x})\ \underline{P}(o)\ (\partial \underline{h}'(\underline{m},o)/\partial \underline{x}) + \underline{R}(o)]\ \underline{\varrho}(o) \tag{2.99}$$

Using the generalized inverse to solve for a $\underline{\varrho}(o)$ and combining (2.96) and (2.99) yields

$$\underline{x}(o) = \underline{m} + \underline{P}(o)\ (\partial \underline{h}'(\underline{m},o)/\partial \underline{x})\ [(\partial \underline{h}(\underline{m},o)/\partial \underline{x})\ \underline{P}(o)$$

$$(\partial \underline{h}'(\underline{m},o)/\partial \underline{x}) + \underline{R}(o)\ ]^{\#}\ \ [\underline{z}(o)-\underline{h}(\underline{m},o)] \tag{2.100}$$

Note that (2.93) and (2.100) are not the same, nor can they be shown to be equal using matrix identities even if $\underline{R}^{-1}(o)$ exists. The reason for this situation is that if $\underline{R}^{-1}$ does not exist we are forced to linearize earlier than if $\underline{R}^{-1}$ exists. The procedure we have used for singular $\underline{R}$ is equivalent to approximating the second term in (2.88) by the following expression;

$$\tfrac{1}{2}\ \| \underline{z}(o)-\underline{h}(\underline{m},o)-(\partial \underline{h}(\underline{m},o)/\partial \underline{x})\ \ [\underline{x}(o)-\underline{m}]\|^2_{\underline{R}^{-1}(o)}$$

Here the approximation takes place before the multiplication; whereas, previously we were able to multiply and then approximate the resulting expression.

The basic structure of the situations which arise when $\underline{h}$ is non-

linear is well illustrated by the simple case n=o. The required modifications of our previous results are evident in (2.91), (2.93) and (2.100).

Suppose that $\underline{w}$ entered the system nonlinearly. We would then approximate

$$\underline{x}(k+1) = \underline{f}(\underline{x}(k), \underline{w}(k), k) \tag{2.3}$$

by

$$\underline{x}(k+1) \approx \underline{f}(\underline{x}^*(k|k),\underline{o},k) + (\partial\, \underline{f}(\underline{x}^*(k|k),\underline{o},k)/\partial\, \underline{x}) \left[\underline{x}(k)-\underline{x}^*(k|k)\right] + (\partial\, \underline{f}(\underline{x}^*(k|k),\underline{o},k)/\partial\, \underline{w})\, \underline{w}(k) \tag{2.101}$$

It is clear that $(\partial\, \underline{f}(\underline{x}^*(k|k),\underline{o},k)/\partial\, \underline{w})$ will play the role of $\underline{G}(k)$ in our previous work.

The reader should have no difficulty in combining these results to meet the needs of any particular problem. How well these approximation techniques will work will, of course, depend on how well the system is approximated, the frequency of measurements, noise levels and other factors peculiar to the individual problem. Similar linearization schemes have been developed from another point of view in recent independent studies by other investigators [37, 59] who give encouraging results.

2.83 Linearity in Disguise

In certain special cases the approximation of (2.1) given by (2.66) becomes exact and equations given by the linearization technique are strictly correct. We rewrite (2.1) and (2.66) for convenience.

$$\underline{x}(k+1) = \underline{f}(\underline{x}(k),k) + \underline{G}(k)\underline{w}(k) \tag{2.1}$$

$$\underline{x}(k+1) = \underline{f}(\underline{x}^*(k|k),k) + (\partial\, \underline{f}(\underline{x}^*(k|k),k)/\partial\, \underline{x}) \left[\underline{x}(k)-\underline{x}^*(k|k)\right] + \underline{G}(k)\underline{w}(k) \tag{2.66}$$

The first case is the trivial one in which $\underline{f}(\underline{x}(k),k)$ is linear. That is, when (2.1) takes the following form;

$$\underline{x}(k+1) = \underline{F}(k)\underline{x}(k) + \underline{G}(k)\underline{w}(k) \tag{2.31}$$

The second case arises when all state variables are exactly measurable and $\underline{x}^*(k|k)$ is equal to $\underline{x}(k)$.

Various combinations of these two cases may arise in which it is not so obvious that the underlying estimation problem is linear even though the basic equations are nonlinear. Consider the case in which the state vector $\underline{x}$ may be decomposed into an exactly observable part $\underline{x}_1$ and a remaining part $\underline{x}_2$. The underlying estimation problem will be linear if the basic system is of the following form;

$$\begin{bmatrix} \underline{x}_1(k+1) \\ \underline{x}_2(k+1) \end{bmatrix} = \begin{bmatrix} \underline{F}_{11}(\underline{x}_1(k),k) & \underline{F}_{12}(\underline{x}_1(k),k) \\ \underline{0} & \underline{F}_{22}(k) \end{bmatrix} \begin{bmatrix} \underline{x}_1(k) \\ \underline{x}_2(k) \end{bmatrix} + \begin{bmatrix} \underline{G}_1(k) \\ \underline{G}_2(k) \end{bmatrix} \underline{w}(k) \tag{2.102}$$

$$\underline{z}(k) = \underline{x}_1(k) \tag{2.103}$$

In this case $\underline{z}$ may be substituted in the equation for $\underline{x}_1$ to produce the following relation;

$$\underline{z}(k+1) - \underline{F}_{11}(\underline{z}(k),k)\ \underline{z}(k) = \underline{F}_{12}(\underline{z}(k),k)\ \underline{x}_2(k) + \underline{G}_1(k)\ \underline{w}(k) \tag{2.104}$$

The relation shows that a difference of known quantities is equal to a linear combination of the unknown part of the state vector plus additive Gaussian noise. This, together with the following linear equation for $\underline{x}_2$

$$\underline{x}_2(k+1) = \underline{F}_{22}(k)\ \underline{x}_2(k) + \underline{G}_2(k)\ \underline{w}(k) \tag{2.105}$$

shows that the estimation problem is linear and the distribution for $\underline{x}_2$

remains Gaussian even though the basic system (2.102) is nonlinear.

For example, consider the following simple system with a random parameter x_2;

$$x_1(k+1) = [\, a + x_2(k) \,]\; x_1(k) + w_1(k)$$
$$x_2(k+1) = b\, x_2(k) + w_2(k)$$
$$z(k) = x_1(k) + v(k)$$

If the additive noise v is not identically zero the estimation problem is nonlinear and the use of the linearization technique would give only an approximate solution to the estimation problem. If v is identically zero this is a special case of (2.102). Then the estimation problem is linear and the linearization technique gives the exact solution to the estimation problem. We see from this simple example that the validity of the approximation may decrease as noise levels increase. This is in addition to the expected decrease in the performance of even the optimal estimator when noise levels increase.

As a second example, consider the following system in which the transition matrix $\underline{F}(k)$ is random;

$$\underline{x}(k+1) = \underline{F}(k)\underline{x}(k) + \underline{G}_1(k)\underline{w}(k) \tag{2.106}$$

$$\underline{z}(k) = \underline{x}(k) \tag{2.107}$$

The elements of the transition matrix $\underline{F}(k)$ are considered to be the outputs of a linear system excited by an independent noise sequence. In order to see how such a situation would be handled we shall examine the following second order system;

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} f_{11}(k) & f_{12}(k) \\ f_{21}(k) & f_{22}(k) \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \underline{G}_1(k)\underline{w}(k)$$

$$z_1(k) = x_1(k) \qquad\qquad z_2(k) = x_2(k)$$

Let us augment the state-space by adjoining the random parameters to the state vector as follows;

$$\begin{bmatrix} f_{11}(k+1) \\ f_{12}(k+1) \\ f_{21}(k+1) \\ f_{22}(k+1) \end{bmatrix} = \begin{bmatrix} x_3(k+1) \\ x_4(k+1) \\ x_5(k+1) \\ x_6(k+1) \end{bmatrix} = \underline{A}(k) \begin{bmatrix} x_3(k) \\ x_4(k) \\ x_5(k) \\ x_6(k) \end{bmatrix} + \underline{G}_2(k)\underline{w}(k)$$

Then the equation for the augmented system may be written in the form of (2.102) as follows;

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \hline x_3(k+1) \\ x_4(k+1) \\ x_5(k+1) \\ x_6(k+1) \end{bmatrix} = \left[\begin{array}{c|cccc} \underline{0} & x_1(k) & x_2(k) & 0 & 0 \\ & 0 & 0 & x_1(k) & x_2(k) \\ \hline \underline{0} & & \underline{A}(k) & & \end{array}\right] \begin{bmatrix} x_1(k) \\ x_2(k) \\ \hline x_3(k) \\ x_4(k) \\ x_5(k) \\ x_6(k) \end{bmatrix} + \begin{bmatrix} \underline{G}_1(k) \\ \hline \underline{G}_2(k) \end{bmatrix} \underline{w}(k)$$

The basic technique used for the second order example is readily extended to the nth order case.

The fact that a nonlinear estimation problem may become linear if certain noise levels go to zero can be a useful tool in investigating nonlinear estimation problems. If we observe that such is the case for a particular estimation problem, we immediately acquire an upper bound on the performance that can be obtained. In addition, we may gain

insight into the coupling between various noises and the quantities to be estimated. For example, in (2.104) we see that the random input $\underline{w}$ is playing the role of additive measurement noise in the problem of estimating $\underline{x}_2$.

Lest we become complacent, consider what would happen if for (2.102) we were able to observe only $\underline{x}_2$. It is evident that we would acquire no new information about $\underline{x}_1$. This difficulty raises the interesting question of observability which will be discussed in Chapter IV.

2.84 Successive Linearization

If one is interested in a problem of ex post facto data analysis the job of obtaining a numerical solution of the two-point boundary value problem arises. As was mentioned earlier, a logical first approximation to the solution would be $\underline{x}^*(k|n)$. One method that might be considered is that of successive linearization. This method is recommended on the grounds that it is comparatively simple and that rapid convergence can be expected. Unfortunately, convergence cannot be guaranteed. This is the price to be paid for simplicity. In a problem where there is no requirement for real time solution it may be worthwhile to try this method first in hopes of simply obtaining a fast solution while running a slight risk of non-convergence. The results of some encouraging applications of this technique are given in Chapter V.

The idea behind the method of successive linearization is this. The estimate $\underline{x}^*(k+1|k+1)$ is generated by linearizing about the point $\underline{x}^*(k|k)$. After the entire observation sequence $\{\underline{z}(o),\dots,\underline{z}(n)\}$ has occurred, the smoothed sequence $\{\underline{x}^*(k|n)\}$, being based on more measurements, would hopefully be a better trajectory about which to linearize than

$\{\underline{x}^*(k|k)\}$ was. One can then use a modified version of (2.77) obtained by linearizing about the sequence $\{\underline{x}^*(k|n)\}$ to reprocess the observations and obtain a sequence $\{\underline{x}_2(k|k)\}$. This sequence should be better than $\{\underline{x}^*(k|k)\}$, since the linearization took place along a better trajectory. This sequence is then smoothed using modified versions of (2.76) and (2.78) to obtain a second approximation to the solution $\{\underline{x}_2(k|n)\}$. The procedure is then repeated linearizing about $\{\underline{x}_2(k|n)\}$.

In the development of the equations for this technique we shall designate the sequence $\{\underline{x}^*(k|n)\}$ by $\{\underline{x}_1(k|n)\}$ to emphasize that it is the first approximation to the solution. Suppose that we have obtained the ith approximation sequence $\underline{x}_i(k|n)$ and we seek approximation i+1. Consider the approximation to (2.1) given by the following linearization;

$$\underline{x}(k+1) \approx \underline{f}(\underline{x}_i(k|n),k) + (\partial \underline{f}(\underline{x}_i(k|n),k)/\partial \underline{x}) \; [\underline{x}(k)-\underline{x}_i(k|n)] + \underline{G}(k)\underline{w}(k) \tag{2.108}$$

If this expression is substituted into (2.9) a modified two-point boundary value problem is obtained;

$$\underline{x}_{i+1}(k+1|n) = \underline{f}(\underline{x}_i(k|n),k) + (\partial \underline{f}(\underline{x}_i(k|n),k)/\partial \underline{x}) \; [\underline{x}_{i+1}(k|n)-\underline{x}_i(k|n)] + \underline{G}(k)\underline{Q}(k)\underline{G}'(k)\underline{\lambda}(k) \tag{2.109}$$

$$\underline{\lambda}(k-1) = (\partial \underline{f}'(\underline{x}_i(k|n),k)/\partial \underline{x}) \; \underline{\lambda}(k) + \underline{H}'(k)\underline{R}^{-1}(k) \; [\underline{z}(k)-\underline{H}(k)\underline{x}_{i+1}(k|n)] \tag{2.110}$$

The boundary conditions are

$$\underline{\lambda}(n) = \underline{o} \tag{2.111}$$

$$\underline{x}_{i+1}(o|n) = \underline{m} + \underline{P}(o)\underline{H}'(o)\underline{R}^{-1}(o) \; [\underline{z}(o)-\underline{H}(o)\underline{x}_{i+1}(o|n)] + \underline{P}(o) \; (\partial \underline{f}'(\underline{x}_i(o|n),o)/\partial \underline{x}) \; \underline{\lambda}(o) \tag{2.112}$$

For simplicity we first assume that the observations are linear and that $\underline{R}(k)$ is positive definite. From (2.109) we see that if it converges, the successive linearization technique converges to a solution of the nonlinear two-point boundary value problem.

The boundary condition (2.112) may be rewritten in the following form;

$$\underline{x}_{i+1}(o|n) = \underline{x}_{i+1}(o|o) + \underline{C}_i(o)(\partial \underline{f}'(\underline{x}_i(o|n),o)/\partial \underline{x})\ \underline{\lambda}(o) \tag{2.113}$$

where $\underline{x}_{i+1}(o|o)$ and $\underline{C}_i(o)$ are $\hat{\underline{x}}(o|o)$ and $\underline{C}(o)$ respectively as given by (2.39) and (2.40).

As usual, we assume a solution of the following form;

$$\underline{x}_{i+1}(k|n) = \underline{x}_{i+1}(k|k) + \underline{C}_i(k)(\partial \underline{f}'(\underline{x}_i(k|n),k)/\partial \underline{x})\ \underline{\lambda}(k) \tag{2.114}$$

Substituting (2.114) into (2.109) yields

$$\underline{x}_{i+1}(k+1|n) = \underline{x}_{i+1}(k+1|k) + \underline{P}_i(k+1)\ \underline{\lambda}(k) \tag{2.115}$$

where $\underline{x}_{i+1}(k+1|k)$ and $\underline{P}_i(k+1)$ are defined as follows;

$$\underline{x}_{i+1}(k+1|k) = \underline{f}(\underline{x}_i(k|n),k) + (\partial \underline{f}(\underline{x}_i(k|n),k)/\partial \underline{x}) \left[\underline{x}_{i+1}(k|k) - \underline{x}_i(k|n)\right] \tag{2.116}$$

$$\underline{P}_i(k+1) = (\partial \underline{f}(\underline{x}_i(k|n),k)/\partial \underline{x})\ \underline{C}_i(k)\ (\partial \underline{f}'(\underline{x}_i(k|n),k)/\partial \underline{x}) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k) \tag{2.117}$$

Combining (2.115) and (2.110) yields

$$\underline{x}_{i+1}(k+1|n) = \underline{x}_{i+1}(k+1|k) + \underline{P}_i(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)\,[\,\underline{z}(k+1) - \underline{H}(k+1)\underline{x}_{i+1}(k+1|n)\,]$$

$$+ \underline{P}_i(k+1)(\partial\,\underline{f}'(\underline{x}_i(k+1|n),k+1)/\partial\,\underline{x})\,\underline{\lambda}(k+1) \qquad (2.118)$$

Solving (2.118) for $\underline{x}_{i+1}(k+1|n)$ we obtain the following relation which satisfies hypothesis (2.114);

$$\underline{x}_{i+1}(k+1|n) = \underline{x}_{i+1}(k+1|k+1) + \underline{C}_i(k+1)(\partial\,\underline{f}'(\underline{x}_i(k+1|n),k+1)/\partial\,\underline{x})\,\underline{\lambda}(k+1) \qquad (2.119)$$

where

$$\underline{x}_{i+1}(k+1|k+1) = \underline{x}_{i+1}(k+1|k) + \underline{C}_i(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)$$

$$[\,\underline{z}(k+1) - \underline{H}(k+1)\underline{x}_{i+1}(k+1|k)\,] \qquad (2.120)$$

and

$$\underline{C}_i(k+1) = [\,\underline{I} + \underline{P}_i(k+1)\underline{H}'(k+1)\underline{R}^{-1}(k+1)\underline{H}(k+1)\,]^{-1}\,\underline{P}_i(k+1) \qquad (2.121)$$

We have now derived a familiar looking set of equations. The sequence $\{\underline{x}_{i+1}(k|k)\}$ may be calculated using (2.116), (2.117), (2.120) and (2.121). The i+1 approximation to the solution $\{\underline{x}_{i+1}(k|n)\}$ may be obtained using (2.120) and (2.110).

These results may be put in an alternate form; combining (2.110), (2.114) and (2.120) yields

$$\underline{\lambda}(k-1) = [\,\underline{I} - \underline{H}'(k)\underline{R}^{-1}(k)\underline{H}(k)\underline{C}_i(k)\,]\;(\partial\,\underline{f}'(\underline{x}_i(k|n),k)/\partial\underline{x})\,\underline{\lambda}(k)$$

$$+ [\,\underline{I} - \underline{H}'(k)\underline{R}^{-1}(k)\underline{H}(k)\underline{C}_i(k)\,]\,\underline{H}'(k)\underline{R}^{-1}(k)\,[\,\underline{z}(k) - \underline{H}(k)\underline{x}_{i+1}(k|k-1)\,] \qquad (2.122)$$

By repeated use of matrix identities this equation may be written in the following form corresponding to (2.57);

$$\underline{\lambda}(k-1) = \left\{ \underline{I} - \underline{H}'(k) \left[\underline{H}(k)\underline{P}_i(k)\underline{H}'(k) + \underline{R}(k)\right]^{-1} \underline{H}(k)\underline{P}(k) \right\}$$

$$(\partial \underline{f}'(\underline{x}_i(k|n),k)/\partial \underline{x})\ \underline{\lambda}(k) + \underline{H}'(k)\ \left[\underline{H}(k)\underline{P}_i(k)\underline{H}'(k) + \underline{R}(k)\right]^{-1}$$

$$\left[\underline{z}(k) - \underline{H}(k)\underline{x}(k|k-1)\right] \tag{2.123}$$

Combining (2.116) and (2.120) and using the matrix identity (A.2) yields

$$\underline{x}_{i+1}(k+1|k) = \underline{f}(\underline{x}_i(k|n),k) + (\partial \underline{f}(\underline{x}_i(k|n),k)/\partial \underline{x})\ \left[\underline{x}_{i+1}(k|k-1) - \underline{x}_i(k|n)\right]$$

$$+ (\partial \underline{f}(\underline{x}_i(k|n),k)/\partial \underline{x})\ \underline{P}_i(k)\ \underline{H}'(k)\ \left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)\right]^{-1}$$

$$\left[\underline{z}(k) - \underline{H}(k)\underline{x}_{i+1}(k|k-1)\right] \tag{2.124}$$

We may also write the following equation for $\underline{P}_i(k)$ by combining (2.117) and (2.121) and using (2.58);

$$\underline{P}_i(k+1) = (\partial \underline{f}(\underline{x}_i(k|n),k)/\partial \underline{x}) \Big\{ \underline{P}_i(k) - \underline{P}_i(k)\underline{H}'(k)\ \left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)\right]^{-1}$$

$$\underline{H}(k)\underline{P}(k)\Big\} (\partial \underline{f}'(\underline{x}_i(k|n),k)/\partial \underline{x}) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k) \tag{2.125}$$

We may summarize the method of obtaining the i+1 approximation sequence $\{\underline{x}_{i+1}(k|n)\}$ from the previous approximation sequence $\{\underline{x}_i(k|n)\}$ in the following three-step procedure.

1. Calculate the sequences $\{\underline{x}_{i+1}(k|k-1)\}$ and $\underline{P}_i(k)$ for $k=o,\ldots,n$ using (2.124) and (2.125) respectively. The procedure begins by setting $\underline{x}_{i+1}(o|-1)$ and $\underline{P}_i(o)$ respectively equal to $\underline{m}$ and $\underline{P}(o)$ of the *a priori* distribution for $\underline{x}(o)$.

2. Calculate the sequence $\{\underline{\lambda}(k)\}$ for $k= -1,\ldots,n$ by backward recursion of (2.123) using the boundary condition $\underline{\lambda}(n) = \underline{o}$.

3. Calculate the i+1 approximation sequence $\{\underline{x}_{i+1}(k|n)\}$ for $k=o,\ldots,n$ using (2.115).

The same technique may be used when $\underline{R}(k)$ is singular if the inverses appearing in (2.123), (2.124) and (2.125) are replaced by generalized inverses.

The method may easily be extended to include the complications of nonlinear observations, correlation between the random input $\underline{w}$ and the measurement noise $\underline{v}$, and the random input $\underline{w}$ entering nonlinearly.

CHAPTER III

ESTIMATION OF STATE VARIABLES FOR CONTINUOUS-TIME SYSTEMS

3.1 Introduction

The mathematical model for the continuous-time estimation problem was discussed in some detail in Chapter I. The explicit problem statement was given in section 1.41.

The continuous-time estimation problem is analogous in many respects to the discrete-time problem of Chapter II. In the continuous-time problem we work with probability density functionals instead of probability density functions. Specifically, we consider

$$p_{x|z}\left[\underline{x}[t_o, t+T] \,\middle|\, \underline{z}[t_o, t]\right]$$

the probability density functional for the time segment of $\underline{x}(t)$ on the interval $[t_o, t+T]$ given the observation $\underline{z}(t)$ on the interval $[t_o, t]$.

An introductory discussion of probability density functionals is given in Appendix E. For the purpose of this chapter it will be sufficient to note that the summations appearing in the discrete-time problem become integrals in the continuous-time problem.

3.2 Preliminary Considerations

We assign a Gaussian *a priori* distribution for the initial state $\underline{x}(t_o)$ with mean $\hat{\underline{x}}(t_o|t_o)$ and covariance matrix $\underline{P}(t_o)$. It will be convenient to assume that $\underline{P}(t_o)$ is positive definite, although this restrictions may be relaxed.

We shall use t to designate the present time, t_o to designate the time observations begin, and τ as a general time parameter. The

estimate of $\underline{x}(\tau)$ given $\underline{z}_{[t_o,t]}$ is denoted by $\hat{\underline{x}}(\tau|t)$.

In order to estimate $\underline{x}_{[t_o,t]}$ when $\underline{G}(t)\underline{Q}(t)\underline{G}'(t)$ is non-singular we may minimize with respect to $\underline{x}_{[t_o,t]}$ the following functional which is the continuous-time analogue of (2.8);

$$J(t) = \tfrac{1}{2}\| \underline{x}(t_o)-\hat{\underline{x}}(t_o|t_o) \|^2_{\underline{P}^{-1}(t_o)} + \tfrac{1}{2}\int_{t_o}^{t} \Big\{ \| \underline{z}(\tau)-\underline{h}(\underline{x}(\tau),\tau) \|^2_{\underline{R}^{-1}(\tau)}$$

$$+ \| \dot{\underline{x}}(\tau)-\underline{f}(\underline{x}(\tau),\tau) \|^2_{[\underline{G}\underline{Q}\underline{G}']^{-1}} \Big\}\, d\tau \qquad (3.1)$$

Equivalently, we may consider minimizing the following functional;

$$I(t) = \tfrac{1}{2}\| \underline{x}(t_o)-\hat{\underline{x}}(t_o|t_o) \|^2_{\underline{P}^{-1}(t_o)}$$

$$+ \tfrac{1}{2}\int_{t_o}^{t} \Big\{ \| \underline{z}(\tau)-\underline{h}(\underline{x}(\tau),\tau) \|^2_{\underline{R}^{-1}(\tau)} + \| \underline{w}(\tau) \|^2_{\underline{Q}^{-1}(\tau)} \Big\}\, d\tau \qquad (3.2)$$

with respect to $\underline{w}_{[t_o,t]}$ and $\underline{x}_{[t_o,t]}$ subject to the constraint

$$\dot{\underline{x}}(\tau) = \underline{f}(\underline{x}(\tau),\tau) + \underline{G}(\tau)\underline{w}(\tau) \qquad (3.3)$$

This functional, (3.2) subject to the constraint (3.3), may also be minimized when $\underline{G}(t)\underline{Q}(t)\underline{G}'(t)$ is singular.

If we wish to estimate $\underline{x}_{[t_o,\ t\,+\,T]}$ (3.2) becomes

$$I(t,t+T) = \frac{1}{2}\|\underline{x}(t_o)-\hat{\underline{x}}(t_o|t_o)\|^2_{\underline{P}^{-1}(t_o)}$$

$$+\frac{1}{2}\int_{t_o}^{t}\left\{\|\underline{z}(\tau)-\underline{h}(\underline{x}(\tau),\tau)\|^2_{\underline{R}^{-1}(\tau)}+\|\underline{w}(\tau)\|^2_{\underline{Q}^{-1}(\tau)}\right\}d\tau$$

$$+\frac{1}{2}\int_{t}^{t+T}\|\underline{w}(\tau)\|^2_{\underline{Q}^{-1}(\tau)}\,d\tau \tag{3.4}$$

That is,

$$I(t,t+T) = I(t) + \frac{1}{2}\int_{t}^{t+T}\|\underline{w}(\tau)\|^2_{\underline{Q}^{-1}(\tau)}\,d\tau \tag{3.5}$$

From (3.5) it is clear that $I(t,t+T)$ may be minimized by first minimizing $I(t)$ and then setting $\underline{w}(\tau)$ equal to zero for $t \leq \tau \leq t+T$. This, together with the constraint (3.3), shows that predictions are again extrapolations of present estimates. We may, therefore, confine our attention to estimating $\underline{x}_{[t_o,t]}$.

3.3 Dynamic Programming Formulation

Note that because of the constraint (3.3), minimizing (3.2) with respect to $\underline{w}_{[t_o,t]}$ and $\underline{x}_{[t_o,t]}$ is equivalent to first minimizing with respect to $\underline{w}_{[t_o,t]}$ and then minimizing with respect to $\underline{x}(t)$. This leads us to define the following "cost" function:

$$V(\underline{x}(t),t) = \underset{\underline{w}_{[t_o,t]}}{\mathrm{Min}}\left\{\frac{1}{2}\|\underline{x}(t_o)-\hat{\underline{x}}(t_o|t_o)\|^2_{\underline{P}^{-1}(t_o)}\right.$$

$$\left.+\frac{1}{2}\int_{t_o}^{t}\left[\|\underline{z}(\tau)-\underline{h}(\underline{x}(\tau),\tau)\|^2_{\underline{R}^{-1}(\tau)}+\|\underline{w}(\tau)\|^2_{\underline{Q}^{-1}(\tau)}\right]d\tau\right\} \tag{3.6}$$

subject to the constraint (3.3).

Note that minimizing with respect to $\underline{w}_{[t_o,t]}$ leaves a boundary condition on $\underline{x}$ to be specified. We satisfy this boundary condition by specifying the final point $\underline{x}(t)$. This constitutes a departure from most dynamic programming problems in which the initial condition $\underline{x}(t_o)$ is specified. Our procedure here is similar to that used in case i of section 2.5, in which f^{-1} existed. No similar assumption need be made here because integration of a differential equation may proceed in either direction, whereas there is a definite direction associated with a difference equation.

We may give $V(\underline{x}(t),t)$ an interpretation similar to the one given to $V_n(\underline{x}(n))$ in the discrete-time problem. Let $\underline{c}$ be any particular value of $\underline{x}(t)$. Then $V(\underline{c},t)$ is a measure of the unlikeliness of the most probable state trajectory $\underline{x}_{[t_o,t]}$ in which $\underline{x}(t)$ takes on the particular value $\underline{c}$, given the observation $\underline{z}_{[t_o,t]}$ and the a priori distribution for $\underline{x}(t_o)$. The estimate of $\underline{x}(t)$ is that value of $\underline{x}(t)$ for which $V(\underline{x}(t),t)$ is minimum.

Using the separability property we may rewrite (3.6) in the following form;

$$V(\underline{x}(t),t) = \underset{\underline{w}_{[t-\Delta t,t]}}{\text{Min}} \left\{ V(\underline{x}(t)-\Delta\underline{x}, t-\Delta t) + \frac{1}{2}\int_{t-\Delta t}^{t} \left[\| \underline{z}(\tau)-\underline{h}(\underline{x}(\tau),\tau) \|^2_{\underline{R}^{-1}(\tau)} + \| \underline{w}(\tau) \|^2_{\underline{Q}^{-1}(\tau)} \right] d\tau \right\}$$

subject to the constraint (3.3). Expanding $V(\underline{x}(t),t)$ in a Taylor series, letting Δt approach zero and using the constraint (3.3), we obtain the following equation;

$$V_t = \underset{\underline{w}(t)}{\text{Min}} \left\{ - V'_{\underline{x}} \, [\, \underline{f}(\underline{x}(t),t) + \underline{G}(t)\underline{w}(t)] \quad + \quad \tfrac{1}{2} \| \underline{z}(t) - \underline{h}(\underline{x}(t),t) \|^2_{\underline{R}^{-1}(t)} \right.$$

$$\left. + \ \tfrac{1}{2} \| \underline{w}(t) \|^2_{\underline{Q}^{-1}(t)} \right\} \tag{3.7}$$

The procedure used to obtain (3.7) is standard in the dynamic programming literature. For more detailed discussions of this procedure and its relation to the calculus of variations see [3, 5, 15, 16].

The $\underline{w}(t)$ which minimizes (3.7) is easily found by differentiation to be

$$\underline{w}(t) = \underline{Q}(t)\underline{G}'(t)V_{\underline{x}} \tag{3.8}$$

Substituting this expression for $\underline{w}(t)$ into (3.7) yields the following important relation;

$$V_t = - \tfrac{1}{2} V'_{\underline{x}} \, \underline{G}(t)\underline{Q}(t)\underline{G}'(t)V_{\underline{x}} - V'_{\underline{x}} \, \underline{f}(\underline{x}(t),t) + \tfrac{1}{2} \| \underline{z}(t) - \underline{h}(\underline{x}(t),t) \|^2_{\underline{R}^{-1}(t)} \tag{3.9}$$

At this point our problem is twofold;

i) Find a function $V(\underline{x}(t),t)$ which satisfies (3.9),

ii) Find $\hat{\underline{x}}(t|t)$, the value of $\underline{x}(t)$ for which $V(\underline{x}(t),t)$ is minimum.

A boundary condition for (3.9) is given by the a priori distribution for the initial state $\underline{x}(t_o)$, since for $t = t_o$, (3.6) becomes

$$V(\underline{x}(t_o),t_o) = \tfrac{1}{2} \| \underline{x}(t_o) - \hat{\underline{x}}(t_o|t_o) \|^2_{\underline{P}^{-1}(t_o)} \tag{3.10}$$

Once $V(\underline{x}(t),t)$ has been found, we may obtain $\hat{\underline{x}}(\tau|t)$ for $t_o \leq \tau \leq t$ by backward integration of the following equation formed by combining

(3.3) and (3.8);

$$\partial \underline{\hat{x}}(\tau|t)/\partial\tau = \underline{f}(\underline{\hat{x}}(\tau|t),\tau) + \underline{G}(\tau)\underline{Q}(\tau)\underline{G}'(\tau)V_{\underline{x}} \qquad (3.11)$$

using $\underline{\hat{x}}(t|t)$ as the boundary condition.

There is no difficulty in including in the dynamic programming formulation systems in which the random input $\underline{w}$ enters nonlinearly. For, example, if the basic system is described by the following equation;

$$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t),t) + \underline{G}(\underline{x}(t),t)\underline{w}(t) \qquad (3.12)$$

then (3.8) becomes

$$\underline{w}(t) = \underline{Q}(t)\underline{G}'(\underline{x}(t),t)V_{\underline{x}} \qquad (3.13)$$

and (3.9) becomes

$$V_t = -\tfrac{1}{2}\, V_{\underline{x}}'\, \underline{G}(\underline{x}(t),t)\underline{Q}(t)\underline{G}'(\underline{x}(t),t)V_{\underline{x}} - V_{\underline{x}}'\, \underline{f}(\underline{x}(t),t)$$

$$+ \tfrac{1}{2}\,\|\underline{z}(t)-\underline{h}(\underline{x}(t),t)\|^2_{\underline{R}^{-1}(t)} \qquad (3.14)$$

For well defined systems described by the following equation;

$$\dot{\underline{x}} = \underline{f}(\underline{x}(t),\underline{w}(t),t) \qquad (3.15)$$

(3.7) becomes

$$V_t = \underset{\underline{w}(t)}{\mathrm{Min}} \left\{ - V_{\underline{x}}'\, \underline{f}(\underline{x}(t),\underline{w}(t),t) + \tfrac{1}{2}\,\|\underline{z}(t)-\underline{h}(\underline{x}(t),t)\|^2_{\underline{R}^{-1}(t)} \right.$$

$$\left. + \tfrac{1}{2}\,\|\underline{w}(t)\|^2_{\underline{Q}^{-1}(t)} \right\} \qquad (3.16)$$

and (3.8) becomes

$$\underline{w}(t) = \underline{Q}(t)(\partial\, \underline{f}'(\underline{x}(t),\underline{w}(t),t)/\partial\, \underline{w})V_{\underline{x}} \tag{3.17}$$

Equation (3.16) is equivalent to two equations: (3.17) and the following equation;

$$V_t = - V_{\underline{x}}' \ \underline{f}(\underline{x}(t),\underline{w}(t),t) + \tfrac{1}{2} \| \underline{z}(t)-\underline{h}(\underline{x}(t),t) \|^2_{\underline{R}^{-1}(t)}$$

$$+ \tfrac{1}{2} \| \underline{w}(t) \|^2_{\underline{Q}^{-1}(t)} \tag{3.18}$$

We shall later relate (3.15) through (3.18) to the two-point boundary value problem obtained by Bryson and Frazier [10].

3.4 Solution to the Linear Problem

The solution of the linear filtering problem discussed in section 1.44 was first obtained by Kalman and Bucy [34]. In this section we give a simplified derivation of this solution based on the dynamic programming formulation of the problem. In order to simplify the resulting equations it will be convenient to omit the parameter t when no confusion is likely to result.

When the basic system is linear (3.9) becomes

$$V_t = - \tfrac{1}{2} \ V_{\underline{x}}' \ \underline{G}(t)\underline{Q}(t)\underline{G}'(t) \ V_{\underline{x}} - V_{\underline{x}}' \ \underline{F}(t)\underline{x}(t) + \tfrac{1}{2} \| \underline{z}(t)-\underline{H}(t)\underline{x}(t) \|^2_{\underline{R}^{-1}(t)} \tag{3.19}$$

It is well known [44, 45] that an analytic solution may be obtained for equations of this type by expressing $V(\underline{x}(t),t)$ in a series of the following form;

$$V(\underline{x}(t),t) = a(t) + \underline{b}'(t)\underline{x}(t) + \tfrac{1}{2}\,\underline{x}'(t)\underline{C}(t)\underline{x}(t)$$

For our purpose it will be more convenient to write this expression in the following symmetric form in which the minimum of the function is shown explicitly;

$$V(\underline{x}(t),t) = \tfrac{1}{2}\,\| \underline{x}(t)-\hat{\underline{x}}(t|t) \|^2_{\underline{P}^{-1}(t)} + k(t) \tag{3.20}$$

In order to determine $\hat{\underline{x}}(t|t)$, $\underline{P}^{-1}(t)$ and $k(t)$ we substitute $V_{\underline{x}}$ and V_t obtained from (3.20) into (3.19) and equate terms of like degree in $\underline{x}(t)$.

Differentiating (3.20) we obtain the following equations;

$$V_{\underline{x}} = \underline{P}^{-1}(t)\,[\,\underline{x}(t)-\hat{\underline{x}}(t|t)] \tag{3.21}$$

$$V_t = \tfrac{1}{2}\,\underline{x}'\,\mathring{\underline{P}}^{-1}\underline{x} - \left[\mathring{\hat{\underline{x}}}'\,\underline{P}^{-1} + \hat{\underline{x}}'\,\mathring{\underline{P}}^{-1}\right]\underline{x} + d(\tfrac{1}{2}\,\hat{\underline{x}}'\,\underline{P}^{-1}\hat{\underline{x}} + k)/dt \tag{3.22}$$

Substituting (3.21) and (3.22) into (3.19) and equating terms of like degree in $\underline{x}$, we obtain the following equations after some algebraic manipulations;

$$\mathring{\underline{P}}^{-1} = -\,\underline{P}^{-1}\underline{GQG}'\,\underline{P}^{-1} - \underline{P}^{-1}\underline{F} - \underline{F}'\underline{P}^{-1} + \underline{H}'\underline{R}^{-1}\underline{H} \tag{3.23}$$

$$d\hat{\underline{x}}(t|t)/dt = \underline{F}(t)\hat{\underline{x}}(t|t) + \underline{P}(t)\underline{H}'(t)\underline{R}^{-1}(t)\,[\,\underline{z}(t)-\underline{H}(t)\hat{\underline{x}}(t|t)\,] \tag{3.24}$$

Using the identity $\mathring{\underline{P}} = -\,\underline{P}\,\mathring{\underline{P}}^{-1}\underline{P}$ we may write (3.23) in the following form;

$$\mathring{\underline{P}}(t) = \underline{F}(t)\underline{P}(t) + \underline{P}(t)\underline{F}'(t) - \underline{P}(t)\underline{H}'(t)\underline{R}^{-1}(t)\underline{H}(t)\underline{P}(t)$$

$$+\,\underline{G}(t)\underline{Q}(t)\underline{G}'(t) \tag{3.25}$$

The final form of the optimal linear filter is given by (3.24)

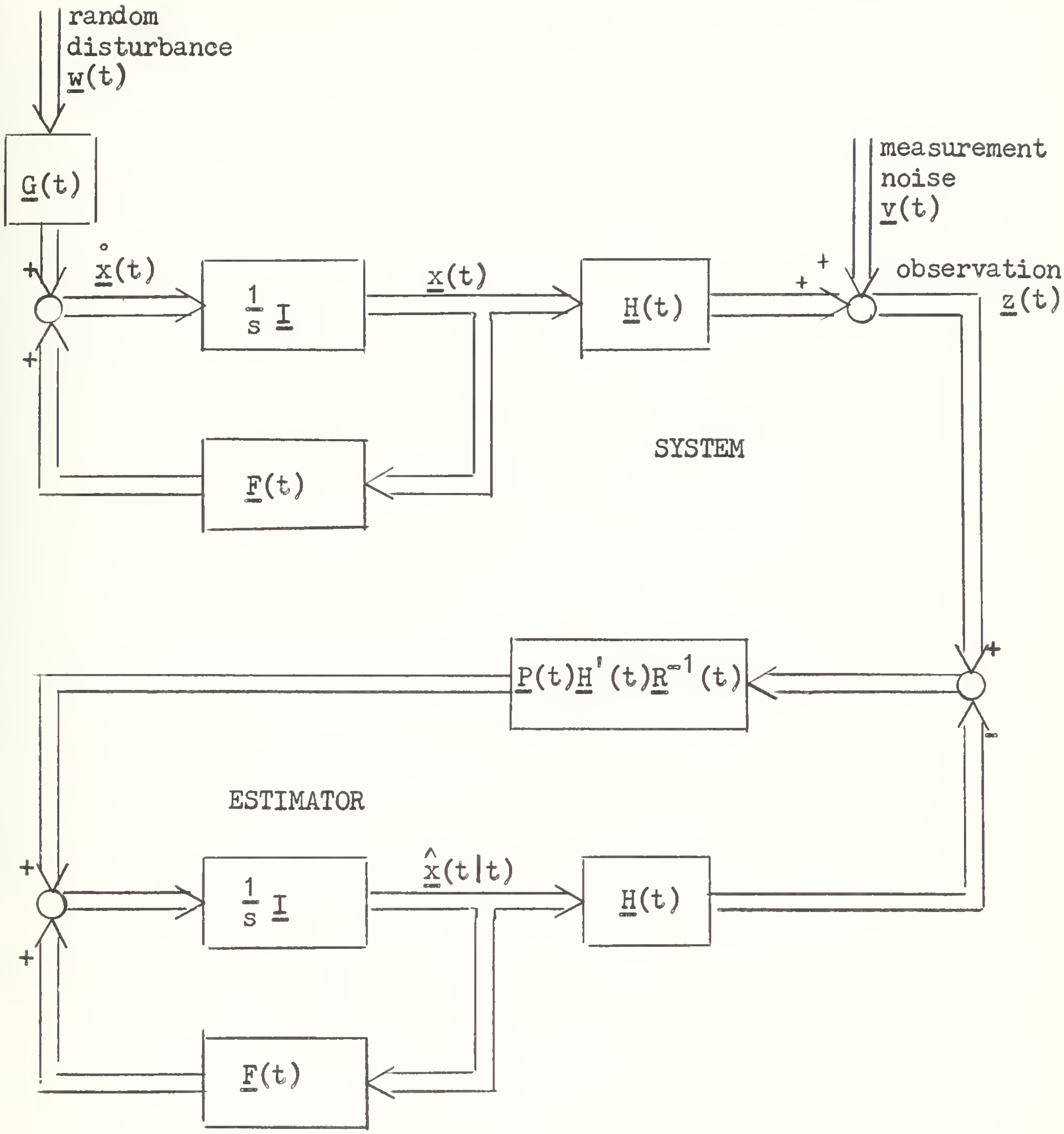


Fig. 10 Optimal linear estimator

and (3.25) which are identical to the equations obtained by Kalman and Bucy. A block diagram of the optimal estimator is given in Fig. 10. Here we again find the intuitively appealing feature of a model of the system appearing in a feedback loop.

As in the discrete-time case, $\underline{P}(t)$ is the covariance matrix of the probability distribution for $\underline{x}(t)$, given the observations. Equations (3.24) and (3.25) remain valid when $\underline{P}(t_o)$ is singular as can be shown by taking the limit of the discrete-time solution. If $\underline{P}(t_o)$ is singular the dynamic programming formulation breaks down because trajectories passing through points which are inconsistent with prior information become infinitely more unlikely than trajectories which do not, and we are unable to reflect this situation in a scalar "cost" function. This is not a serious drawback since the a priori distribution may be chosen to approximate the state of prior knowledge arbitrarily closely without introducing a singular covariance matrix. As in the discrete-time case, this difficulty is largely bypassed if $\underline{GQG}'$ is non-singular, since then $\underline{P}(t_o + \Delta t)$ is positive definite.

3.5 Approximate Solution to the Nonlinear Problem

In order to develop an approximate solution to the nonlinear problem we follow a procedure similar to the one used in discrete-time. We may rewrite (3.9) in the following form;

$$V_t = -\tfrac{1}{2}\left\| \underline{G}'(t)V_{\underline{x}} \right\|^2_{\underline{Q}(t)} - V_{\underline{x}}' \; \underline{f}(\underline{x}(t),t) + \tfrac{1}{2}\left\| \underline{z}(t)-\underline{h}(\underline{x}^*(t|t),t) \right.$$

$$\left. - \quad \underline{h}(\underline{x}(t),t)-\underline{h}(\underline{x}^*(t|t),t) \right\|^2_{\underline{R}^{-1}(t)} \tag{3.26}$$

Since the estimate of $\underline{x}(t)$ is the value of $\underline{x}(t)$ for which $V(\underline{x}(t),t)$ is minimum, we would expect that $V(\underline{x}(t),t)$ could be closely approximated in a neighborhood of this minimum by a quadratic form in $\underline{x}$. Such a quadratic form has already been shown to be the solution in the linear case. We shall use $\underline{x}^*(t|t)$ to designate the estimate produced by this approximation. The terms on the right hand side of (3.26) involving $\underline{f}$ and $\underline{h}$ will be expanded in a Taylor series about the point $\underline{x}^*(t|t)$. The argument t should be understood but will usually not be written. All partial derivatives are to be evaluated at the point $(\underline{x}^*(t|t),t)$.

The approximation for $\underline{f}$ is simply

$$\underline{f}(\underline{x}(t),t) \approx \underline{f}(\underline{x}^*) + (\partial \underline{f}/\partial \underline{x})[\ \underline{x}-\underline{x}^*\ ] \tag{3.27}$$

The term involving $\underline{h}$ may be broken down into the following sum of three terms;

$$\tfrac{1}{2}\,\|\underline{z}-\underline{h}(\underline{x}^*)\|^2_{\underline{R}^{-1}} + \tfrac{1}{2}\,\|\underline{h}(\underline{x})-\underline{h}(\underline{x}^*)\|^2_{\underline{R}^{-1}} - [\underline{h}'(\underline{x})-\underline{h}'(\underline{x}^*)]\ \underline{R}^{-1}\,[\underline{z}-\underline{h}(\underline{x}^*)]$$

The first of these terms is independent of $\underline{x}$. The second term is approximated as follows;

$$\tfrac{1}{2}\,\|\underline{h}(\underline{x})-\underline{h}(\underline{x}^*)\|^2_{\underline{R}^{-1}} \approx \tfrac{1}{2}\,[\underline{x}-\underline{x}^*\ ]'\ \left\{\partial[(\partial\underline{h}'/\partial\underline{x})\underline{R}^{-1}\underline{h}]/\partial\underline{x}\right\}[\underline{x}-\underline{x}^*] \tag{3.28}$$

The approximation for the third term is

$$-[\underline{h}'(\underline{x})-\underline{h}'(\underline{x}^*)]\ \underline{R}^{-1}\ [\underline{z}-\underline{h}(\underline{x}^*)] \approx -[\underline{x}-\underline{x}^*]'\ (\partial\underline{h}'/\partial\underline{x})\ \underline{R}^{-1}\,[\underline{z}-\underline{h}(\underline{x}^*)] \tag{3.29}$$

If these approximations are substituted into (3.26) the following equation results;

$$V_t = -\tfrac{1}{2}\| \underline{G}'V_{\underline{x}} \|^2_{\underline{Q}} - V'_{\underline{x}} \left\{ \underline{f}(\underline{x}^*) + (\partial \underline{f}/\partial \underline{x}) \; [\underline{x}-\underline{x}^*] \right\}$$

$$+ \tfrac{1}{2}\| \underline{z}-\underline{h}(\underline{x}^*) \|^2_{\underline{R}^{-1}} + \tfrac{1}{2}[\underline{x}-\underline{x}^*]' \; \left\{ \partial[(\partial \underline{h}'/\partial \underline{x})\underline{R}^{-1}\underline{h}] \,/\partial \underline{x} \right\} [\underline{x}-\underline{x}^*]$$

$$- [\underline{x}-\underline{x}^*]' \; (\partial \underline{h}'/\partial \underline{x})\underline{R}^{-1} \; [\underline{z}-\underline{h}(\underline{x}^*)] \qquad (3.30)$$

An analytic solution of (3.30) may be obtained by assuming that $V(\underline{x}(t),t)$ is quadratic in $\underline{x}$. Specifically,

$$V(\underline{x}(t),t) = \tfrac{1}{2}\| \underline{x}(t)-\underline{x}^*(t|t) \|^2_{\underline{P}^{*-1}(t)} + k(t) \qquad (3.31)$$

It is straightforward, but algebraically involved, to substitute into (3.30) the $V_{\underline{x}}$ and V_t obtained from (3.31) to obtain the following equations which describe the estimator;

$$d\underline{x}^*(t|t)/dt = \underline{f}(\underline{x}^*(t|t),t) + \underline{P}^*(t)(\partial \, \underline{h}'(\underline{x}^*(t|t),t)/\partial \underline{x})\underline{R}^{-1}(t)$$

$$[\underline{z}(t)-\underline{h}(\underline{x}^*(t|t),t)] \qquad (3.32)$$

$$\mathring{\underline{P}}^*(t) = (\partial \underline{f}/\partial \underline{x})\underline{P}^*(t) + \underline{P}^*(t)(\partial \underline{f}'/\partial \underline{x}) + \underline{G}(t)\underline{Q}(t)\underline{G}'(t)$$

$$- \underline{P}^*(t) \; \left\{ \partial[(\partial \underline{h}'/\partial \underline{x})\underline{R}^{-1}\underline{h}\,]/\partial \underline{x} \right\} \underline{P}^*(t) \qquad (3.33)$$

The approximate solution to the problem of estimating the state variables of a nonlinear system in real time is given by (3.32) and (3.33). For linear systems these equations reduce to those obtained by Kalman and Bucy. The initial conditions for $\underline{P}^*(t)$ and $\underline{x}^*(t|t)$ are the $\underline{P}(t_o)$ and $\hat{\underline{x}}(t_o|t_o)$ given by the a priori distribution for the initial state $\underline{x}(t_o)$.

A block diagram of the nonlinear estimator in a control system

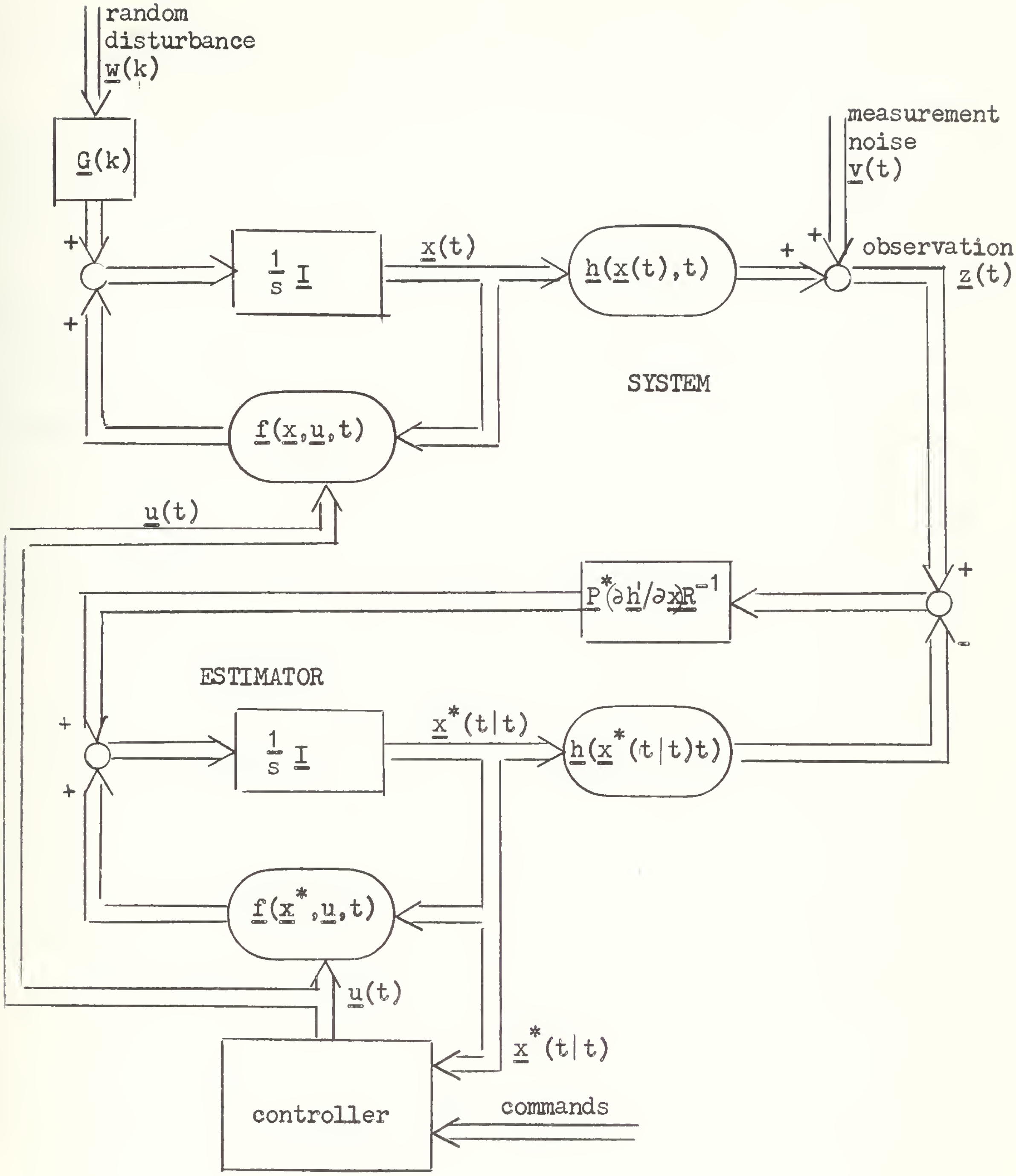


Fig. 11 Nonlinear estimator in a control system

is given in Fig. 11. In the diagram the control vector $\underline{u}$ is shown explicitly. Again we find the intuitively appealing design which places a model of the basic system in a feedback loop.

If the random input $\underline{w}$ enters the system nonlinearly we may retain the basic structure of the estimator if we replace $\underline{G}(t)$ by $\underline{G}(\underline{x}^*(t|t),t)$ or $(\partial \underline{f}(\underline{x}^*(t|t),\underline{0},t)/\partial \underline{w})$ -- whichever is appropriate.

A first approximation to the solution of the problem of estimating $\underline{x}(\tau)$ for $t_0 \leq \tau \leq t$ is given by $\underline{x}^*(\tau|t)$ which can be obtained by backward integration of the following equation using $\underline{x}^*(t|t)$ as the boundary condition;

$$\partial \underline{x}^*(\tau|t)/\partial \tau = \underline{f}(\underline{x}^*(\tau|t),\tau) + \underline{G}(\tau)\underline{Q}(\tau)\underline{G}'(\tau)V_{\underline{x}} \tag{3.34}$$

3.6 A Two-Point Boundary Value Problem

The two-point boundary value problem obtained by Bryson and Frazier [10] using the calculus of variations follows directly from the dynamic programming point of view. We shall consider the most general case in which the noise enters nonlinearly as described by (3.15) through (3.18). The procedure we follow is based on the one used by Dreyfus [16] in his study of the problem of Mayer in the calculus of variations.

In order to simplify the resulting expressions the arguments $\underline{x}$, $\underline{w}$ and t usually will be omitted when no confusion is likely to result. The following relation follows from the rules of differentiation;

$$dV_{\underline{x}}/dt = \partial V_{\underline{x}}/\partial t + (\partial V_{\underline{x}}/\partial \underline{x})\, \dot{\underline{x}} \tag{3.35}$$

Differentiating (3.18) with respect to $\underline{x}$ yields

$$\partial \underline{V}_{\underline{x}}/\partial t = -(\partial \underline{f}'/\partial \underline{x})\underline{V}_{\underline{x}} - (\partial \underline{V}_{\underline{x}}/\partial \underline{x})\overset{\circ}{\underline{x}}$$

$$-(\partial \underline{h}'/\partial \underline{x})\ \underline{R}^{-1}\ [\ \underline{z}-\underline{h}(\underline{x},t)]$$

$$-(\partial \underline{w}'/\partial \underline{x})\ [\ (\partial \underline{f}'/\partial \underline{w})\ V_{\underline{x}} - \underline{Q}^{-1}\underline{w}\ ] \tag{3.36}$$

Noting that by (3.17) the coefficient of $(\partial \underline{w}'/\partial \underline{x})$ in (3.36) is zero, we may combine (3.35) and (3.36) to obtain the following equation;

$$dV_{\underline{x}}/dt = -(\partial \underline{f}'/\partial \underline{x})V_{\underline{x}} - (\partial \underline{h}'/\partial \underline{x})\ \underline{R}^{-1}\ [\underline{z}-\underline{h}(\underline{x},t)] \tag{3.37}$$

This equation, together with (3.15) and (3.17), are the basic relations for the two-point boundary value problem. At the final time t we have the following natural boundary condition which corresponds to the requirement that $\hat{\underline{x}}(t|t)$ is the value of $\underline{x}(t)$ for which $V(\underline{x}(t),t)$ is minimum;

$$V_{\underline{x}}(t) = \underline{0} \tag{3.38}$$

The boundary condition at the initial time t_o is obtained by differentiating (3.10). This yields

$$\hat{\underline{x}}(t_o|t) = \hat{\underline{x}}(t_o|t_o) + \underline{P}(t_o)V_{\underline{x}} \tag{3.39}$$

The quantity $V_{\underline{x}}$ plays the role of the Lagrange multiplier of the calculus of variations [15, 16]. If we let $V_{\underline{x}} = \underline{\lambda}(t)$, then the two-point boundary value problem may be written in the following form which closely resembles that obtained by Bryson and Frazier;

$$\partial \hat{\underline{x}}(\tau|t)/\partial\tau = \underline{f}(\hat{\underline{x}}(\tau|t),\underline{w}(\tau),\tau) \qquad t_o \leq \tau \leq t \tag{3.40}$$

$$\underline{w}(\tau) = \underline{Q}(\tau)(\partial \underline{f}'/\partial \underline{w})\ \underline{\lambda}(\tau) \qquad t_o \leq \tau \leq t \tag{3.41}$$

$$\overset{\circ}{\underline{\lambda}}(\tau) = -(\partial \underline{f}'/\partial \underline{x})\, \underline{\lambda}(\tau) - (\partial \underline{h}'/\partial \underline{x})\, \underline{R}^{-1}(\tau)\, [\underline{z}(\tau) - \underline{h}(\hat{\underline{x}}(\tau|t),\tau)]$$

$$t_o \leq \tau \leq t \qquad (3.42)$$

The boundary conditions are

$$\underline{\lambda}(t) = \underline{o} \qquad (3.43)$$

$$\hat{\underline{x}}(t_o|t) = \hat{\underline{x}}(t_o|t_o) + \underline{P}(t_o)\, \underline{\lambda}(t_o) \qquad (3.44)$$

For the important special cases in which the random input $\underline{w}$ enters linearly, the equations for $\hat{\underline{x}}(\tau|t)$ and $\underline{\lambda}(\tau)$ become

$$\partial \hat{\underline{x}}(\tau|t)/\partial\tau = \underline{f}(\hat{\underline{x}}(\tau|t),\tau) + \underline{G}(\tau)\underline{Q}(\tau)\underline{G}'(\tau)\, \underline{\lambda}(\tau) \qquad (3.45)$$

$$\overset{\circ}{\underline{\lambda}}(\tau) = -(\partial \underline{f}'/\partial \underline{x})\, \underline{\lambda}(\tau) - (\partial \underline{h}'/\partial \underline{x})\, \underline{R}^{-1}(\tau)\, [\underline{z}(\tau) - \underline{h}(\hat{\underline{x}}(\tau|t),\tau)]$$

$$(3.46)$$

We shall have more to say about the two-point boundary value problem later when we discuss the relationship between the estimation problem and control problems.

3.7 Continuous-Time Systems with Sampled Observations

In certain applications observations of a continuous-time system can conveniently be made at discrete points in time. We shall discuss this situation only briefly, since it requires only minor modifications of our previous results.

Consider the system

$$\overset{\circ}{\underline{x}}(t) = \underline{f}(\underline{x}(t),t) + \underline{G}(t)\underline{w}(t) \qquad (3.47)$$

Suppose the observations are made at equally spaced intervals of length T. The observed signal is

$$\underline{z}(kT) = \underline{h}(\underline{x}(kT), kT) + \underline{v}(kT) \tag{3.48}$$

We again assume that the random input $\underline{w}(t)$ and the measurement noise $\underline{v}(kT)$ are Gaussian with zero mean. The covariance matrices for $\underline{v}$ and $\underline{w}$ are as follows;

$$\begin{aligned} E[\underline{w}(t)\underline{w}'(\tau)] &= \underline{Q}(t) \; \delta(t-\tau) \\ E[\underline{v}(kT)\underline{v}'(jT)] &= \underline{R}(kT) \quad \delta_{jk} \\ E[\underline{w}(t)\underline{v}'(kT)] &= \underline{0} \end{aligned}$$

If we are between observations n and n+1 ($nT \leq t < (n+1)T$), then by analogy to previous work, we may minimize the following functional;

$$I(t) = \tfrac{1}{2} \| \underline{x}(o) - \underline{\hat{x}}(o|o^-) \|^2_{\underline{P}^{-1}(o)} + \int_o^t \tfrac{1}{2} \| \underline{w}(\tau) \|^2_{\underline{Q}^{-1}(\tau)} \, d\tau$$

$$+ \sum_{k=o}^{n} \tfrac{1}{2} \| \underline{z}(kT) - \underline{h}(\underline{x}(kT), kT) \|^2_{\underline{R}^{-1}(kT)} \qquad nT \leq t < (n+1)T \tag{3.49}$$

subject to the constraint (3.47).

Here we have assumed for convenience that observations begin at t=o. Note that because of the sampled nature of the observations, the estimate will be discontinuous at the sampling times. We use $\underline{\hat{x}}(o|o^-)$ to designate the mean of the a priori distribution for the initial state. This corresponds to $\underline{m}$ of the discrete-time problem. Similarly, $\underline{\hat{x}}(o|o^+)$ would correspond to $\underline{\hat{x}}(o|o)$ of the discrete-time problem.

Using a standard engineering technique, we let i(t) be a train of unit impulses with the impulses occurring at the sampling times kT. Then (3.49) may be rewritten in the following form which closely resembles (3.2);

$$I(t) = \tfrac{1}{2} \| \underline{x}(o) - \hat{\underline{x}}(o|o^-) \|^2_{\underline{P}^{-1}(o)} + \int_0^t \left\{ i(\tau) \| \underline{z}(\tau) - \underline{h}(\underline{x}(\tau),\tau) \|^2_{\underline{R}^{-1}(\tau)} + \| \underline{w}(\tau) \|^2_{\underline{Q}^{-1}(\tau)} \right\} d\tau \qquad (3.50)$$

subject to the constrant (3.47). Proceeding formally

$$V_t = - \tfrac{1}{2} V_{\underline{x}}' \underline{GQG}' V_{\underline{x}} - V_{\underline{x}}' \underline{f}(\underline{x}(t),t) + \tfrac{1}{2} i(t) \| \underline{z}(t) - \underline{h}(\underline{x}(t),t) \|^2_{\underline{R}^{-1}(t)} \qquad (3.51)$$

Note that V_t is discontinuous at the sampling times so that the limit used to obtain (3.51) by analogy to (3.7) does not exist at the sampling instants. Proceeding in a purely formal manner, we now use the same approximation technique that was employed in section 3.5 to obtain the following equations;

$$d\underline{x}^*(t|t)/dt = \underline{f}(\underline{x}^*(t|t),t) + i(t)\, \underline{P}^*(t^+)\, (\partial \underline{h}'/\partial \underline{x}) \underline{R}^{-1} [\, \underline{z}(t) - \underline{h}(\underline{x}^*(t|t),t)] \qquad (3.52)$$

$$\mathring{\underline{P}}^{*-1} = - \underline{P}^{*-1} (\partial \underline{f}/\partial \underline{x}) - (\partial \underline{f}'/\partial \underline{x})\, \underline{P}^{*-1} - \underline{P}^{*-1}\, \underline{GQG}'\, \underline{P}^{*-1} + i(t) \left\{ \partial [(\partial \underline{h}'/\partial \underline{x})\, \underline{R}^{-1}\, \underline{h}\,] / \partial \underline{x} \right\} \qquad (3.53)$$

At the sampling instants the identity $\mathring{\underline{P}}^{*-1} = - \underline{P}^* \mathring{\underline{P}}^{*-1} \underline{P}^*$ breaks down. Between the sampling times we have

$$\mathring{\underline{P}}^*(t) = (\partial \underline{f}/\partial \underline{x})\, \underline{P}^*(t) + \underline{P}^*(t)\, (\partial \underline{f}'/\partial \underline{x}) + \underline{G}(t)\underline{Q}(t)\underline{G}'(t) \quad \text{for } kT < t < (k+1)T \qquad (3.54)$$

At the sampling instants we have the following formula resembling the results of the discrete-time problem;

$$\underline{P}^*(kT^+) = \left[\underline{I} + \underline{P}(kT^-) \left\{ \partial [(\partial \underline{h}'/\partial \underline{x})\, \underline{R}^{-1}\, \underline{h}\,] / \partial \underline{x} \right\} \right]^{-1} \underline{P}(kT^-) \qquad (3.55)$$

Note that the values of $\underline{P}^*(kT^+)$ and $(\partial \underline{h}/\partial \underline{x})$ are only needed at the sampling times. The estimator could, therefore

consist of an analogue model of the basic system in conjunction with a special purpose digital computer. Such a hybrid scheme would simultaneously avoid the difficulties of approximating the nonlinear system on a digital computer and performing a large number of multiplications on an analogue computer.

3.8 Relation Between the Estimation Problem and Control Problems

The fact that the solution of the linear filtering problem is formally the same as the solution of the problem of controlling a linear system with quadratic performance criteria was first pointed out by Kalman [31] and the existence of this duality is by now well known. Since this duality has received ample coverage in the literature [25, 29, 31, 32, 33, 34] we shall not dwell on this point except to mention that the nature of this duality is made especially clear by the present formulation of the estimation problem in which the estimation procedure involves the minimization of a quadratic "cost" functional.

The close connection between the observability of a system and the controllability of the adjoint system will be discussed in Chapter IV.

It should be clear from the present formulation that there is a close connection between the nonlinear estimation problem and certain nonlinear control problems. For example, consider the following system:

$$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t),t) + \underline{G}(t)\underline{u}(t)$$
$$\underline{y}(t) = \underline{h}(\underline{x}(t),t)$$

Suppose that we want the output $\underline{y}(t)$ to follow a known function $\underline{z}(t)$ and agree to choose the control $\underline{u}_{[t_o,T]}$ to minimize the following performance functional:

$$\frac{1}{2} \int_{t_0}^{T} \left[\| \underline{z}(t) - \underline{h}(\underline{x}(t),t) \|^2_{\underline{R}^{-1}(t)} + \| \underline{u}(t) \|^2_{\underline{Q}^{-1}(t)} \right] dt$$

This functional is a weighted combination of a quadratic measure of the error and the total "energy" expended.

Various methods may be used to attack this problem. We shall find it convenient to use the Hamiltonian point of view, so much in vogue in the current control literature. Since this section is an aside, we shall assume that the reader is familiar with the basic concepts. For extensive treatment of these and related matters, the reader is referred to [3, 5, 15, 16, 30, 52, 55, 61].

Let us define the following Lagrangian;

$$L(\underline{x}(t),\underline{u}(t),t) = \tfrac{1}{2} \| \underline{z}(t) - \underline{h}(\underline{x}(t),t) \|^2_{\underline{R}^{-1}(t)} + \| \underline{u}(t) \|^2_{\underline{Q}^{-1}(t)}$$

Using the "Maximum Principle", we consider the following Hamiltonian;

$$H(\underline{x}(t),\underline{p}(t),t) = \underset{\underline{u}(t)}{\mathrm{Max}} \left\{ \underline{p}'(t) \left[\underline{f}(\underline{x}(t),t) + \underline{G}(t)\underline{u}(t) \right] - L(\underline{x}(t),\underline{u}(t),t) \right\}$$

The maximizing value of $\underline{u}(t)$ is easily obtained by differentiation

$$\underline{u}^*(t) = \underline{Q}(t)\underline{G}'(t)\underline{p}(t)$$

Substituting this value into the expression for H yields

$$H(\underline{x}(t),\underline{p}(t),t) = \underline{p}'(t)\underline{f}(\underline{x}(t),t) + \tfrac{1}{2} \| \underline{G}'(t)\underline{p}(t) \|^2_{\underline{Q}} - \tfrac{1}{2} \| \underline{z}(t) - \underline{h}(\underline{x}(t),t) \|^2_{\underline{R}^{-1}(t)}$$

The canonical equations for this problem,

$$\dot{\underline{x}}(t) = \partial H / \partial \underline{p}$$

$$\dot{\underline{p}}(t) = - \partial H / \partial \underline{x}$$

are the same as (3.45) and (3.46) if we equate $\underline{p}(t)$ and $\underline{\lambda}(t)$. Further, we can easily recognize (3.9) as the Hamilton-Jacobi partial differential equation [15, 16, 30],

$$V_t + H(\underline{x}(t), V_{\underline{x}}, t) = o$$

We can make the boundary conditions the same as those for the estimation problem by leaving the final state $\underline{x}(T)$ unspecified and by assigning a quadratic "cost" $\| \underline{x}(t_o) - \underline{x}^*(t_o) \|^2_{\underline{P}^{-1}(t_o)}$ for starting in a state other than a nominal initial state $\underline{x}^*(t_o)$. Specifying the initial state would correspond to the smoothing problem in which the initial state were known exactly; that is, $\underline{P}(t_o) = \underline{0}$.

The regulator problem $\underline{z}(t) \equiv \underline{o}$ corresponds to a smoothing problem in which the observed value of signal plus noise is equal to zero on $[t_o, T]$. While such a set of observations is extremely unlikely, it is, nevertheless, the most likely curve $\underline{z}_{[t_o, T]}$ for a stable system with the mean of the a priori distribution for $\underline{x}(t_o)$ equal to zero.

While this formulation of the control problem is somewhat artificial since, except for the regulator problem, the desired output is usually not known in advance, the point we wish to make is that the problem of finding the minimum energy control so that the output of the system follows a desired trajectory using a quadratic performance criteria is mathematically identical with the problem of finding the most likely random input after making noisy observations of the output.

CHAPTER IV

OBSERVABILITY AND CONTROLLABILITY

4.1 Introduction

The purpose of this chapter is to give a brief introduction to the concepts of observability and controllability and to relate these ideas to the present formulation of the estimation problem. These concepts were introduced by Kalman [32, 34] and have been studied by him [33] and his associates [70]. The paper by Bertram and Sarachik represents an early contribution, [69].

We shall confine our attention to discrete-time systems since certain problems arise in discrete-time systems which do not exist in their continuous-time counterparts and since the results are easily extended to continuous-time systems which have been studied more extensively in the recent literature [33, 70].

The definitions of controllability and observability which we shall introduce will differ slightly from those given elsewhere. The modified definitions will enable us to include linear systems with singular transition matrices and -- in principle -- nonlinear systems. They will also simplify the development of the relation between observability of a system and controllability of the adjoint system which is of particular significance in the estimation problem.

The material in section B6 of Appendix B will be continually referred to and provides necessary background for this chapter.

4.2 Preliminary Considerations

Consider the following linear discrete-time system;

$$\underline{x}(k+1) = \underline{F}(k)\underline{x}(k) + \underline{G}(k)\underline{u}(k) \tag{4.1}$$

$$\underline{y}(k) = \underline{H}(k)\underline{x}(k) \tag{4.2}$$

In section 1.3 we showed that the forward transition matrix for this system may be expressed in the following form;

$$\underline{\Phi}(m+1,k+1) = \underline{F}(m)\underline{F}(m-1) \circ \circ \circ \underline{F}(k+1) \qquad m > k \tag{4.3}$$

The adjoint system of (4.1) and (4.2) is defined to be

$$\underline{x}(k-1) = \underline{F}'(k)\underline{x}(k) + \underline{H}'(k)\underline{u}(k) \tag{4.4}$$

$$\underline{y}(k) = \underline{G}'(k)\underline{x}(k) \tag{4.5}$$

The backward transition matrix for the adjoint system is

$$\underline{\Psi}(k,m) = \underline{F}'(k+1) \circ \circ \circ \underline{F}'(m-1)\underline{F}'(m) \qquad m > k \tag{4.6}$$

Comparing (4.3) and (4.6) we obtain the following important identity;

$$\underline{\Phi}(m+1,k+1) = \underline{\Psi}'(k,m) \tag{4.7}$$

The relation between a system and its adjoint will be discussed in some detail in the sequel.

4.3 Definition of Observability

The question of observability has to do with the basic structure of the system under consideration. Roughly, the question is this; "Given a finite sequence of exact measurements of the output of an unforced system, can one uniquely reconstruct the corresponding state sequence?" If the answer is yes, the system will be called completely observable on the sequence of times during which the measurements are made.

In order to develop some insight we shall first examine the question of observability for a time-invariant discrete-time system before giving a more precise definition of observability and considering its ramifications. Consider the following system;

$$\underline{x}(k+1) = \underline{F}\,\underline{x}(k) + \underline{G}\,\underline{u}(k) \tag{4.8}$$

$$\underline{y}(k) = \underline{H}\,\underline{x}(k) \tag{4.9}$$

with the state vector $\underline{x}$, and n-vector, and the output $\underline{y}$. If the system is unforced, $\underline{u} \equiv \underline{o}$. Suppose that we make n consecutive measurements of the output $\underline{y}$. For convenience we assume that the first measurement occurs at time zero. Clearly, this involves no loss of generality in the time-invariant case. We may then write the following equations relating the output sequence $\{\underline{y}(o), \ldots, \underline{y}(n-1)\}$ and the initial state $\underline{x}(o)$;

$$\begin{aligned}
\underline{y}(o) &= \underline{H}\,\underline{x}(o) \\
\underline{y}(1) &= \underline{H}\,\underline{F}\,\underline{x}(o) \\
&\vdots \\
\underline{y}(n-1) &= \underline{H}\,\underline{F}^{n-1}\,\underline{x}(o)
\end{aligned}$$

or, equivalently

$$\begin{bmatrix} \underline{y}(o) \\ \underline{y}(1) \\ \vdots \\ \underline{y}(n-1) \end{bmatrix} = \begin{bmatrix} \underline{H} \\ \underline{H}\,\underline{F} \\ \vdots \\ \underline{H}\,\underline{F}^{n-1} \end{bmatrix} \underline{x}(o) \tag{4.10}$$

We wish to know whether or not the initial state $\underline{x}(o)$ can be determined

uniquely as a result of the measurements of the output. Once the initial state $\underline{x}(o)$ is known, the sequence of states $\{\underline{x}(k)\}$ (for $k \geq o$) of the unforced system is given by (4.8). We implicitly assume that (4.10) has a solution; that is the output sequence does not contradict the model of (4.8) and (4.9). From elementary matrix theory we know that $\underline{x}(o)$ can be determined uniquely if and only if the matrix on the right side of (4.10) or, equivalently, its transpose $[\underline{H}', \underline{F}'\underline{H}', \ldots, \underline{F}'^{n-1}\underline{H}']$ has rank n. Moreover, if the initial state of the time-invariant system cannot be determined after n observations of the output, additional observations will convey no new information because, by the Cayley-Hamilton theorem [17], any positive integer power of an n by n matrix F may be expressed as a linear combination of I, F, ..., F^{n-1} and, therefore, additional observations are simply linear combinations of the first n observations.

In devising a workable definition of observability for discrete-time systems we are faced with a problem that does not arise in continuous-time systems, namely that a discrete-time system may be described by a forward difference equation, for example (4.8); or by a backward difference equation, for example (2.34).

Consider the forward system

$$\underline{x}(k+1) = \underline{f}(\underline{x}(k),k) \tag{4.11}$$

and the backward system

$$\underline{x}(k-1) = \underline{f}(\underline{x}(k),k) \tag{4.12}$$

the output of both systems being given by an equation of the following form;

$$\underline{y}(k) = \underline{h}(\underline{x}(k),k) \tag{4.13}$$

In order to include both types of systems, we propose the following definition of observability.

<u>Definition</u>: A discrete-time system will be called <u>completely observable on a set of times</u> $\{j, j+1, \ldots, m-1, m\}$ if, by observing the output sequence $\{\underline{y}(j), \ldots, \underline{y}(m)\}$, one can uniquely determine the state sequence $\{\underline{x}(j), \ldots, \underline{x}(m)\}$ for all possible state sequences.

It is clear that for the forward system (4.11) and (4.13) it is necessary and sufficient to determine $\underline{x}(j)$ and for the backward system (4.12) and (4.13) it is necessary and sufficient to determine $\underline{x}(m)$.

4.4 Definition of Controllability

The question of controllability, like the question of observability, has to do with the basic structure of the system under consideration. Roughly, the question is this; "Given any two states $\underline{x}_1$ and $\underline{x}_2$, can one choose a finite input sequence in such a manner that the corresponding state sequence will be of the form $\{\underline{x}_1, \ldots, \underline{x}_2\}$?" If the answer is yes, the system will be called completely controllable on the set of times during which the state sequence occurs.

In order to develop some insight we shall again consider the question of controllability for a time-invariant discrete-time system before giving a more precise definition of controllability. Consider the following system;

$$\underline{x}(k+1) = \underline{F}\,\underline{x}(k) + \underline{G}\,\underline{u}(k) \tag{4.14}$$

with the state vector $\underline{x}$, an n-vector, and the control input, $\underline{u}$. Suppose that we wish to move the system from a known initial state $\underline{x}(o)$ to a desired state $\underline{x}_d$ at time n. We may then write the following equations

relating the state sequence and the control sequence;

$$\underline{x}(1) = \underline{F}\,\underline{x}(o) + \underline{G}\,\underline{u}(o)$$

$$\underline{x}(2) = \underline{F}^2\,\underline{x}(o) + \underline{F}\,\underline{G}\,\underline{u}(o) + \underline{G}\,\underline{u}(1)$$

$$\vdots$$

$$\underline{x}_d(n) = \underline{F}^n\,\underline{x}(o) + \underline{F}^{n-1}\,\underline{G}\,\underline{u}(o) + \cdots + \underline{G}\,\underline{u}(n-1) \tag{4.15}$$

We may rewrite this last equation in the following form;

$$\underline{x}_d(n) - \underline{F}^n\,\underline{x}(o) = [\,\underline{F}^{n-1}\,\underline{G},\ldots,\underline{F}\,\underline{G},\underline{G}\,] \begin{bmatrix} \underline{u}(o) \\ \vdots \\ \underline{u}(n-1) \end{bmatrix} \tag{4.16}$$

We see that a solution of (4.16) will exist if and only if $\underline{x}_d(n) - \underline{F}^n\,\underline{x}(o)$ is a linear combination of the column vectors of the matrix $[\,\underline{F}^{n-1}\,\underline{G},\ldots,\underline{F}\,\underline{G},\underline{G}\,]$; in which case $\underline{x}_d(n) - \underline{F}^n\,\underline{x}(o)$ belongs to the range of that matrix. If the matrix $[\,\underline{F}^{n-1}\,\underline{G},\ldots,\underline{F}\,\underline{G},\underline{G}\,]$ has rank n, then any final state may be reached from any initial state in n transitions. Moreover, by the Cayley-Hamilton theorem, any state which can be reached from $\underline{x}(o)$ can be reached in (at most) n transitions. For the special case of a single input system, the matrix $\underline{G}$ becomes a vector $\underline{g}$ and we obtain the results that every final state may be reached from every initial state if and only if the matrix $[\underline{F}^{n-1}\,\underline{g},\ldots,\underline{F}\,\underline{g},\underline{g}]$ is non-singular and that any state which can be reached, can be reached in (at most) n transitions. These results are well known [32, 69] and are the basis for the design of "dead beat" systems.

In devising a workable definition of controllability we are again faced with the problem of forward and backward difference equations. Consider

the forward system;

$$\underline{x}(k+1) = \underline{f}(\underline{x}(k), \underline{u}(k), k) \tag{4.17}$$

and the backward system;

$$\underline{x}(k-1) = \underline{f}(\underline{x}(k), \underline{u}(k), k) \tag{4.18}$$

In order to include both types of systems, we propose the following definition of controllability.

Definition: A discrete-time system will be called completley controllable on a set of times $\{j, j+1, \ldots, m-1, m\}$ if for every $\underline{x}_1$ and $\underline{x}_2$ a control sequence $\{\underline{u}(j), \ldots, \underline{u}(m)\}$ exists such that $\underline{x}(j) = \underline{x}_1$ and $\underline{x}(m) = \underline{x}_2$.

It is clear that for the forward system (4.17) only the sequence $\{\underline{u}(j), \ldots, \underline{u}(m-1)\}$ is pertinent and for the backward system (4.18) only the sequence $\{\underline{u}(j+1), \ldots, \underline{u}(m)\}$ is pertinent.

4.5 Controllability and Observability for Linear Discrete-Time Systems

The questions of observability and controllability for time-varying linear discrete-time systems are closely related to those for the time-invariant systems already discussed. Let us consider the system of (4.1) and (4.2) which we rewrite for convenience

$$\underline{x}(k+1) = \underline{F}(k)\underline{x}(k) + \underline{G}(k)\underline{u}(k) \tag{4.1}$$

$$\underline{y}(k) = \underline{H}(k)\underline{x}(k) \tag{4.2}$$

If we observe the output sequence $\{\underline{y}(j), \ldots, \underline{y}(m)\}$ for the unforced system, $\underline{u} \equiv 0$, then we may write the following equation relating the output sequence $\{\underline{y}(j), \ldots, \underline{y}(m)\}$ to the initial state $\underline{x}(j)$;

$$\begin{bmatrix} \underline{y}(j) \\ \underline{y}(j+1) \\ \vdots \\ \underline{y}(m) \end{bmatrix} = \begin{bmatrix} \underline{H}(j) \\ \underline{H}(j+1)\,\underline{\Phi}(j+1,j) \\ \vdots \\ \underline{H}(j+m)\,\underline{\Phi}(m,j) \end{bmatrix} \underline{x}(j) \tag{4.19}$$

Since this is just a special case of the equation $\underline{Ax} = \underline{y}$ which is considered in Appendix B, we may apply lemma B6.1 to obtain the following result.

Lemma 4.51: The system (4.1) and (4.2) is completely observable on the set of times $\{j, j+1, \ldots, m-1, m\}$ if and only if the following matrix is positive definite;

$$\underline{M}_f(j,m) = \sum_{k=j}^{m} \underline{\Phi}'(k,j)\underline{H}'(k)\underline{H}(k)\underline{\Phi}(k,j) \tag{4.20}$$

If $\underline{M}_f(j,m)$ is positive definite, then the unique solution for $\underline{x}(j)$ is

$$\underline{x}(j) = \underline{M}_f^{-1}(j,m) \sum_{k=j}^{m} \underline{\Phi}'(k,j)\underline{H}'(k)\underline{y}(k) \tag{4.21}$$

° ° °

From (4.21) we see that it is not necessary to know the sequence $\{\underline{y}(j), \ldots, \underline{y}(m)\}$ but merely the following sum;

$$\sum_{k=j}^{m} \underline{\Phi}'(k,j)\underline{H}'(k)\underline{y}(k)$$

Note that if $\underline{M}_f(j,m)$ is positive definite, then so is $\underline{M}_f(j,m+k)$ for $k > o$.

Let us now consider the question of controllability of the same system, (4.1). If we consider the control sequence $\{\underline{u}(j), \ldots, \underline{u}(m-1)\}$,

then we may write the following equation relating the states $\underline{x}(j)$ and $\underline{x}(m)$ and the control sequence

$$\underline{x}(m) - \underline{\Phi}(m,j)\underline{x}(j) = [\underline{\Phi}(m,j+1)\underline{G}(j),\ldots,\underline{G}(m-1)] \begin{bmatrix} \underline{u}(j) \\ \cdot \\ \cdot \\ \cdot \\ \underline{u}(m-1) \end{bmatrix} \tag{4.22}$$

Again this is just a special case of the equation $\underline{Ax} = \underline{y}$ and we immediately obtain the following lemmas as special cases of lemmas B6.2 and B6.3 respectively.

Lemma 4.52: The state $\underline{x}(m)$ may be reached at time m from the state $\underline{x}(j)$ at time j if and only if $\underline{x}(m) - \underline{\Phi}(m,j)\underline{x}(j)$ belongs to the range of the following matrix;

$$\underline{W}_f(j,m) = \sum_{k=j}^{m-1} \underline{\Phi}(m,k+1)\underline{G}(k)\underline{G}'(k)\underline{\Phi}'(m,k+1) \tag{4.23}$$

In this case, the solution which minimizes

$$\sum_{k=j}^{m-1} \tfrac{1}{2} \|\underline{u}(k)\|^2$$

is

$$\underline{u}(k) = \underline{G}'(k)\underline{\Phi}'(m,k+1)\underline{W}_f^{\#}(j,m)\ [\underline{x}(m) - \underline{\Phi}(m,j)\underline{x}(j)] \qquad k=j,\ldots,m-1 \tag{4.24}$$

° ° °

Lemma 4.53: The system (4.1) is completely controllable on the set of times $\{j,j+1,\ldots,m-1,m\}$ if and only if $\underline{W}_f(j,m)$ is positive definite. In this case the solution which minimizes

$$\sum_{k=j}^{m-1} \tfrac{1}{2} \| \underline{u}(k) \|^2$$

is

$$\underline{u}(k) = \underline{G}'(k)\underline{\Phi}'(m,k+1)\underline{W}_f^{-1}(j,m)\ [\underline{x}(m) - \underline{\Phi}(m,j)\underline{x}(j)] \quad k=j,\ldots,m-1 \quad (4.25)$$

$\circ \quad \circ \quad \circ$

Let us now consider a system described by a backward difference equation. In particular, let us consider the adjoint system given by (4.4) and (4.5) which we rewrite for convenience

$$\underline{x}(k-1) = \underline{F}'(k)\underline{x}(k) + \underline{H}'(k)\underline{u}(k) \quad (4.4)$$

$$\underline{y}(k) = \underline{G}'(k)\underline{x}(k) \quad (4.5)$$

In order to examine the question of observability we write the following equation relating the output sequence $\{\underline{y}(j),\ldots,\underline{y}(m)\}$ and the state $\underline{x}(m)$ for the unforced system;

$$\begin{bmatrix} \underline{y}(j) \\ \circ \\ \circ \\ \circ \\ \underline{y}(m-1) \\ \underline{y}(m) \end{bmatrix} = \begin{bmatrix} \underline{G}'(j)\ \underline{\Psi}(j,m) \\ \circ \\ \circ \\ \circ \\ \underline{G}'(m-1)\ \underline{\Psi}(m-1,m) \\ \underline{G}'(m) \end{bmatrix} \underline{x}(m) \quad (4.26)$$

Again we note that (4.26) is just a special case of the equation $\underline{Ax} = \underline{y}$ and we immediately obtain the following lemma as a special case of B6.1.

Lemma: 4.54: The system (4.4) and (4.5) is completely observable on the set of times $\{j,\ldots,m\}$ if and only if the following matrix is positive definite;

$$\underline{M}_b(j,m) = \sum_{k=j}^{m} \underline{\Psi}'(k,m)\underline{G}(k)\underline{G}'(k)\underline{\Psi}(k,m) \tag{4.27}$$

If $\underline{M}_b(j,m)$ is positive definite the unique solution for $\underline{x}(m)$ is

$$\underline{x}(m) = \underline{M}_b^{-1}(j,m) \sum_{k=j}^{m} \underline{\Psi}'(k,m)\underline{G}(k)\underline{y}(k) \tag{4.28}$$

° ° °

Using the identity (4.7) we may write (4.27) as follows;

$$\underline{M}_b(j,m) = \sum_{k=j}^{m} \underline{\Phi}(m+1,k+1)\underline{G}(k)\underline{G}'(k)\underline{\Phi}'(m+1,k+1) = \underline{W}_f(j,m+1) \tag{4.29}$$

A comparison of (4.23) and (4.29) yields the following theorem.

<u>Theorem 4.55</u>: The system (4.1) and (4.2) is completely controllable on the set of times $\{j,\ldots,m+1\}$ if and only if the adjoint system (4.4) and (4.5) is completely observable on the set of times $\{j,\ldots,m\}$.

° ° °

Let us now examine the question of controllability of the adjoint system. We consider the control sequence $\{\underline{u}(j+1),\ldots,\underline{u}(m)\}$ and obtain the following equation relating the states $\underline{x}(j)$ and $\underline{x}(m)$ and the control sequence;

$$\underline{x}(j) - \underline{\Psi}(j,m)\underline{x}(m) = \left[\underline{H}'(j+1),\ldots,\underline{\Psi}(j,m-2)\underline{H}'(m-1),\underline{\Psi}(j,m-1)\underline{H}'(m)\right] \begin{bmatrix} \underline{u}(j+1) \\ \circ \\ \circ \\ \circ \\ \underline{u}(m-1) \\ \underline{u}(m) \end{bmatrix} \tag{4.30}$$

Again this is just a special case of the equation $\underline{Ax} = \underline{y}$ and we obtain the following lemma as a special case of B6.3.

Lemma 4.56: The system (4.5) is completely controllable on the set of times $\{j, j+1, \ldots, m-1, m\}$ if and only if the following matrix is positive definite;

$$\underline{W}_b(j,m) = \sum_{k=j+1}^{m} \underline{\Phi}(j,k-1)\underline{H}'(k)\underline{H}(k)\underline{\Phi}'(j,k-1) \tag{4.31}$$

If $\underline{W}_b(j,m)$ is positive definite the solution which minimizes

$$\sum_{k=j+1}^{m} \tfrac{1}{2} \|\underline{u}(k)\|^2$$

is

$$\underline{u}(k) = \underline{H}(k)\underline{\Phi}'(j,k-1)\underline{W}_b^{-1}(j,m) \; [\underline{x}(j) - \underline{\Phi}(j,m)\underline{x}(m)] \quad k=j+1,\ldots,m \tag{4.32}$$

∘ ∘ ∘

Using the identity (4.7) we may rewrite (4.31) in the following form;

$$\underline{W}_b(j,m) = \sum_{k=j+1}^{m} \underline{\Phi}'(k,j+1)\underline{H}'(k)\underline{H}(k)\underline{\Phi}(k,j+1) = \underline{M}_f(j+1,m) \tag{4.33}$$

A comparison of (4.20) and (4.33) yields the following theorem.

Theorem 4.57: The system (4.1) and (4.2) is completely observable on the set of times $\{j, j+1, \ldots, m\}$ if and only if the adjoint system (4.5) and (4.6) is completely controllable on the set of times $\{j-1, j, \ldots, m\}$.

∘ ∘ ∘

It is interesting to note that the lemmas given elsewhere [70]

correspond to the results we have obtained for the backward system. For discrete-time systems in which $\underline{F}^{-1}(k)$ exists and for continuous-time systems the distinction between forward and backward equations does not arise. In these cases the basic lemmas may be stated in alternate forms corresponding to those we have obtained for the forward and backward systems.

We may easily relate theorem 4.57 to the general solution of the linear smoothing problem given in Appendix D. Consider the case in which the random input $\underline{w}$ and the measurement noise $\underline{v}$ are uncorrelated. For convenience we rewrite (D.47) in the following form;

$$\underline{\lambda}(k-1) = \underline{F}'(k)\underline{\lambda}(k) + \underline{H}'(k) \left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)\right]^{\#}$$
$$\left\{\underline{z}(k) - \underline{H}(k)\left[\hat{\underline{x}}(k|k-1) + \underline{P}(k)\underline{F}'(k)\underline{\lambda}(k)\right]\right\} \qquad (4.34)$$

In addition from (D.48) we obtain

$$\hat{\underline{x}}(o|n) = \underline{m} + \underline{P}(o)\underline{\lambda}(-1) \qquad (4.35)$$

From (4.35) we see that the estimate of the initial state may take on all possible values only if $\underline{\lambda}(-1)$ can take on all possible values, but from (4.34) we see that $\underline{\lambda}(-1)$ can take on all possible values only if the adjoint system is completely controllable on the set of times $\{-1,o,\ldots,n\}$. The solution of the linear estimation problem given in Appendix D is optimal even if the system is not completely observable and the optimal procedure is simply to estimate the a priori mean for those states which are unobservable.

If we note that the expression

$$\underline{z}(k) - \underline{H}(k)\left[\hat{\underline{x}}(k|k-1) + \underline{P}(k)\underline{F}'(k)\underline{\lambda}(k)\right]$$

appearing in (4.34) is the estimate of the measurement noise $\underline{v}(k)$ we

see that an absurd situation would arise if the system were not completely observable and the adjoint system were completely controllable. For then we would be using the estimate of the measurement noise to modify the *a priori* estimate of certain state variables when the measurements themselves contained no information about those state variables.

We may also relate theorem 4.55 to a simple control problem. Consider the problem of moving the system (4.1) from a specified initial state $\underline{x}(j)$ to a desired state $\underline{x}_d$ in n transitions using the control sequence for which the "energy" expended is minimum. "Energy" is defined to be

$$\sum_{k=j}^{m-1} \tfrac{1}{2} \left\| \underline{u}(k) \right\|^2 \qquad ,m = j+n$$

To solve this problem we minimize the following expression;

$$\sum_{k=o}^{m-1} \tfrac{1}{2} \left\| \underline{u}(k) \right\|^2 + \underline{\lambda}'(k) \left[\underline{x}(k+1) - \underline{F}(k)\underline{x}(k) - \underline{G}(k)\underline{u}(k) \right] \tag{4.36}$$

subject to the constraint

$$\underline{x}_d(m) - \underline{\Phi}(m,j)\underline{x}(j) = \sum_{k=j}^{m-1} \underline{\Phi}(m,k+1)\underline{G}(k)\underline{u}(k) \tag{4.37}$$

Setting the partial derivatives of (4.36) with respect to $\underline{u}(k)$, $\underline{\lambda}(k)$ and $\underline{x}(k)$ equal to zero yields the following equations;

$$\underline{x}(k+1) = \underline{F}(k)\underline{x}(k) + \underline{G}(k)\underline{\lambda}(k) \tag{4.1}$$

$$\underline{\lambda}(k-1) = \underline{F}'(k)\underline{\lambda}(k) \tag{4.38}$$

$$\underline{u}(k) = \underline{G}'(k)\underline{\lambda}(k) \tag{4.39}$$

Note that $\underline{u}(k)$ is the output of the adjoint system. From (4.38) we may obtain the following expression;

$$\underline{\lambda}(k) = \underline{\Phi}(k,m-1)\underline{\lambda}(m-1) \tag{4.40}$$

Substituting (4.38) and (4.40) into (4.37) yields

$$\underline{x}_d(m) - \underline{\Phi}(m,j)\underline{x}(j) = \sum_{k=j}^{m-1} \underline{\Phi}(m,k+1)\underline{G}(k)\underline{G}'(k)\underline{\Phi}(k,m-1)\underline{\lambda}(m-1) \tag{4.41}$$

Using the identity (4.7) we may rewrite (4.41) in the following two equivalent forms;

$$\underline{x}_d(m) - \underline{\Phi}(m,j)\underline{x}(j) = \underline{W}_f(j,m)\underline{\lambda}(m-1) \qquad ,m = j+n \tag{4.42}$$

$$\underline{x}_d(m) - \underline{\Phi}(m,j)\underline{x}(j) = \underline{M}_b(j,m-1)\underline{\lambda}(m-1) \qquad ,m = j+n \tag{4.43}$$

Lemma 4.52 says that the state $\underline{x}_d$ may be reached from $\underline{x}(j)$ in n transitions if and only if some $\underline{\lambda}(m-1)$ satisfies (4.42). Lemma 4.53 says that every state may be reached from $\underline{x}(j)$ in n transitions if and only if there is a unique $\underline{\lambda}(m-1)$ satisfying 4.42. If we consider $\underline{u}(k)$ to be the output of the adjoint system then, by lemma 4.54, $\underline{\lambda}(m-1)$ can be uniquely determined on the basis of the following sum;

$$\underline{x}_d(m) - \underline{\Phi}(m,j)\underline{x}(j) = \sum_{k=j}^{m-1} \underline{\Phi}'(m-1,k)\underline{G}(k)\underline{u}(k)$$

if and only if the adjoint system is completely observable on the set of times $\{j,\ldots,m-1\}$. Theorem 4.56 simply says that the system is completely controllable if and only if the Lagrange multiplier $\underline{\lambda}(m-1)$ may be determined uniquely; that is, if and only if the adjoint system is completely observable on the set of times $\{j,\ldots,m-1\}$.

4.6 Nonlinear Systems

At present almost nothing is known about controllability and observability for nonlinear systems. A theorem on local controllability of nonlinear systems was proven by Markus and Lee [71] by assuming that the linear variational equations were completely controllable. The discussion of their paper by Kalman is also of interest.

In spite of the present lack of understanding of these matters, there still seems to be a strong connection between observability of a nonlinear system and controllability of the adjoint system. This is indicated by the fact that the ubiquitous two-point boundary value problems involve the adjoint equations. For example, consider a system in which the state vector may be divided into a completely unobservable part $\underline{x}_2$ and a remaining part $\underline{x}_1$, in such a way that the equations for the system may be written in the following form;

$$\underline{x}_1(k+1) = \underline{f}_1(\underline{x}_1(k), \underline{u}(k), k)$$
$$\underline{x}_2(k+1) = \underline{f}_2(\underline{x}_1(k), \underline{x}_2(k), \underline{u}(k), k)$$

$$\underline{y}(k) = \underline{h}(\underline{x}_1(k), k)$$

If we examine the adjoint system

$$\underline{\lambda}(k-1) = (\partial \underline{f}'/\partial \underline{x})\, \underline{\lambda}(k) + (\partial \underline{h}'/\partial \underline{x})\, \underline{v}(k)$$

or, more specifically

$$\begin{bmatrix} \underline{\lambda}_1(k-1) \\ \underline{\lambda}_2(k-1) \end{bmatrix} = \begin{bmatrix} \partial \underline{f}_1'/\partial \underline{x}_1 & \partial \underline{f}_2'/\partial \underline{x}_1 \\ \underline{0} & \partial \underline{f}_2'/\partial \underline{x}_2 \end{bmatrix} \begin{bmatrix} \underline{\lambda}_1(k) \\ \underline{\lambda}_2(k) \end{bmatrix} + \begin{bmatrix} \partial \underline{h}'/\partial \underline{x}_1 \\ \underline{0} \end{bmatrix} \underline{v}(k)$$

we find that there is no trajectory along which the linearized adjoint system is completely controllable.

Similarly, if we consider a system with obviously uncontrollable state variables of the following form;

$$\underline{x}_1(k+1) = \underline{f}_1(\underline{x}_1(k),\underline{x}_2(k),\underline{u}(k),k)$$
$$\underline{x}_2(k+1) = \underline{f}_2(\underline{x}_2(k),k)$$

we find that there is no trajectory along which the following adjoint system is completely observable;

$$\begin{bmatrix} \underline{\lambda}_1(k-1) \\ \underline{\lambda}_2(k-1) \end{bmatrix} = \begin{bmatrix} \partial \underline{f}_1'/\partial \underline{x}_1 & \underline{0} \\ \partial \underline{f}_1'/\partial \underline{x}_2 & \partial \underline{f}_2'/\partial \underline{x}_2 \end{bmatrix} \begin{bmatrix} \underline{\lambda}_1(k) \\ \underline{\lambda}_2(k) \end{bmatrix}$$

$$\underline{y} = [\, \partial \underline{f}_1'/\partial \underline{u}' \quad \underline{0} \,] \begin{bmatrix} \underline{\lambda}_1(k) \\ \underline{\lambda}_2(k) \end{bmatrix}$$

If we were to consider the problem of observing a small perturbation of the initial state $\underline{x}(j)$ of the system (4.11) and (4.13) we would investigate the variational equations

$$\delta \underline{x}(k+1) = (\partial \underline{f}/\partial \underline{x})\ \delta \underline{x}(k)$$
$$\delta \underline{y}(k) = (\partial \underline{h}/\partial \underline{x})\ \delta \underline{x}(k)$$

Complete observability of the variational equations along the nominal trajectory would insure local observability of a small perturbation in $\underline{x}(j)$. Complete observability of the variational equations is, of course, the same as complete controllability of the adjoint equations.

CHAPTER V

EXPERIMENTAL RESULTS

5.1 Introduction

In this chapter we shall present the results of the application of several of the techniques of Chapter II to some particular examples. The aim of these numerical studies is to give an indication of the type of performance that can be expected when approximation methods are used in nonlinear estimation problems. With regard to the approximation technique for the filtering problem given in section 2.81; the behavior of the $\underline{P}^*$ matrix, the stability of the filter, the effect of known control inputs, and the overall quality of the estimates, are all of prime importance. With regard to the smoothing problem; the quality of the first approximation sequence $\{\underline{x}^*(k|n)\}$ and the convergence properties of the successive linearization technique, are the basic factors to be considered. While one cannot draw positive general conclusions from specific examples, we feel that the results of these simulation studies indicate that the methods under consideration hold great promise.

5.2 Pictures and Random Variables

The simulation studies discussed in this chapter were performed on an IBM 7090 computer. The results will be displayed in a series of pictures; each of which represents a time history of 1000 points, time running from left to right. The scale for each picture will be indicated by giving the maximum in absolute value of the variable or by referring to an adjacent picture of the same scale. Positive values of the variables lie <u>below</u> the horizontal axis.

Gaussian random variables were approximated by a sum of twelve independent pseudo-random variables, each having a uniform distribution on the interval $[0,1]$.

5.3 Example: As a first example a first order system with a randomly varying parameter was investigated. Such a system could be a model for a noisy RC network in which the resistance varied in a random manner. The continuous-time version of the system is given by the following equations;

$$\dot{x}_1(t) = [a + x_2(t)]\, x_1(t) + u(t) + w_1(t) \tag{5.1}$$

$$\dot{x}_2(t) = b\, x_2(t) + w_2(t) \tag{5.2}$$

$$z(t) = x_1(t) + v(t) \tag{5.3}$$

where $u(t)$ is a known input; $w_1(t)$, $w_2(t)$ and $v(t)$ are white Gaussian noise processes; $z(t)$ is the observed output; and the random parameter $x_2(t)$ is the output of a linear system excited by white Gaussain noise. Because the simulation was to be conducted on a digital computer, this system was approximated by the following discrete-time system;

$$x_1(t+h) = x_1(t) + [a + x_2(t)]\, x_1(t)h + u(t)h + w_1(t)h \tag{5.4}$$

$$x_2(t+h) = x_2(t) + b\, x_2(t)h + w_2(t)h \tag{5.5}$$

$$z(t) = x_1(t) + v(t) \tag{5.6}$$

For $a = -1.0$, the value of .01 was used for h. This corresponds to a sampling period of 10 milliseconds for a system with an average time constant of one second.

The recursive scheme of (2.73), (2.74) and (2.77) was used to obtain estimates of x_1 and x_2 in real time. The successive linearization technique of section 2.84 was used to obtain a solution of the smoothing problem. The equations describing the estimation procedure for this

problem can easily be obtained by straightforward substitution into the basic equations of Chapter II.

5.31 Case I

The following data pertains to the simulation results shown in Plate I.

system parameters:	$a = -1.0$ $b = -0.1$ $h = 0.01$	
initial conditions:	$x_1(o) = 1.0$ $x_2(o) = -0.5$	
a priori distribution:	$m_1 = 0.0$ $m_2 = 0.0$	$p_{11}(o) = 0.5$ $p_{12}(o) = 0.0$ $p_{22}(o) = 0.6$
noise levels:	$E[w_1(j)w_1(k)] = 36.0\ \delta_{jk}$ $E[w_2(j)w_2(k)] = 1.0\ \delta_{jk}$ $E[v(j)v(k)] = 9.0\ \delta_{jk}$	

w_1, w_2 and v are statistically independent.

Discussion: This data set represents a rather severe noise condition as indicated by parts B and C of Plate I. The real time estimate of x_1 closely followed the actual value. The error in the *a priori* estimate of x_2 was overcome soon after the known input u excited a transient, as can be seen in parts G and H.

The results of the application of the successive linearization technique to the smoothing problem are shown in parts F and I. The iteration procedure was terminated when the maximum deviations between the ith and the i+1 approximations for both x_1 and x_2 were less than one percent of the maximum value of the quantity being estimated. This criteria was satisfied by the third approximation. The difference between the

first three approximations could not be detected in the pictures. Note that in the smoothed solution for x_2 shown in Part I, the error in the <u>a priori</u> estimate was overcome.

The known input u was chosen as a representative example of the type of saturating signals occurring in control systems. The total computer time used for the entire simulation was 0.73 minutes.

<u>5.32 Case II</u>

The following data pertains to the simulation results shown in Plate II.

system parameters: $a = -1.0$, $b = 0.0$, $h = 0.01$

initial conditions: $x_1(o) = 2.0$, $x_2(o) = -1.5$

<u>a priori</u> distribution: $m_1 = 0.0$, $m_2 = 0.0$; $p_{11}(o) = 1.0$, $p_{12}(o) = 0.0$, $p_{22}(o) = 0.6$

noise levels:

$$E\left[w_1(j)w_1(k)\right] = 16.0\ \delta_{jk}$$

$$E\left[w_2(j)w_2(k)\right] = 0.0$$

$$E\left[v(j)v(k)\right] = 9.0\ \delta_{jk}$$

w_1, w_2 and v are statistically independent.

<u>Discussion</u>: For this data set the parameter x_2 is an unknown constant. The severity of the noise conditions is indicated by parts B and F of Plate II. In part D it can be seen that soon after the known input u excited a transient, the estimator was able to estimate x_2 very closely. The estimate of x_1 is shown in part H. The improvement in the estimate of x_1 with increasing time can be attributed to the information in the transient and the improved estimate of x_2.

5.4 Example: As a second example, a second order system with a randomly varying natural frequency was investigated. The continuous-time version of the system is given in the example of section 1.43. Because the simulation was to be conducted on a digital computer, this system was approximated by the following discrete-time system;

$$x_1(t+h) = x_1(t) + x_2(t)h \tag{5.7}$$

$$x_2(t+h) = x_2(t) - b\, x_2(t)h - [\, c_o + x_3(t)\,]\, x_1(t)h + u(t)h + w_1(t)h \tag{5.8}$$

$$x_3(t+h) = x_3(t) - a\, x_3(t)h + w_2(t)h \tag{5.9}$$

$$z(t) = x_1(t) + v(t) \tag{5.10}$$

The recursive scheme of (2.73), (2.74) and (2.77) was used to obtain estimates of x_1, x_2 and x_3 in real time. The equations describing the estimation procedure for this problem are easily obtained by straightforward substitution into the basic equations of Chapter II.

5.41 Case III

The following data pertains to the simulation results shown in Plate III.

system parameters:

$$a = 0.1$$
$$b = 0.2$$
$$c_o = 3.0$$
$$h = 0.01$$

initial conditions:

$$x_1(o) = 0.5$$
$$x_2(o) = 0.5$$
$$x_3(o) = -0.5$$

a priori distribution:

$m_1 = 0.0$	$p_{11}(o) = 1.0$	$p_{22}(o) = 0.5$
$m_2 = 0.0$	$p_{12}(o) = 0.0$	$p_{23}(o) = 0.0$
$m_3 = 0.0$	$p_{13}(o) = 0.0$	$p_{33}(o) = 0.5$

noise levels: $E\,[w_1(j)w_1(k)] = 0.16\ \delta_{jk}$

$E\,[w_2(j)w_2(k)] = 225.0\ \delta_{jk}$

$E\,[v(j)v(k)] = 0.64\ \delta_{jk}$

w_1, w_2 and v are statistically independent.

<u>Discussion</u>: In this data set large and rapid changes in the natural frequency $[c_o + x_3]$ occur. The noise conditions are moderate as shown in parts B and C of Plate III. The real time estimates of the state variables are seen to follow satisfactorily the rapid changes in the state variables. Note that, since x_2 is the discrete-time analogy of the derivative of x_1, the job of estimating this quantity is intrinsically difficult. Again there is improved behavior after the known input u excites a transient. The behavior of the $\underline{P}^*$ matrix shown in parts J through O, is typical of the behavior observed in a large number of runs. The total computer time used for the entire simulation was 0.74 minutes.

<u>5.42 Case IV</u>

The following data pertains to the simulation results shown in Plate IV.

system parameters:
$$\begin{aligned} a &= 0.0 \\ b &= 1.0 \\ c_o &= 0.0 \\ h &= 0.01 \end{aligned}$$

initial conditions:
$$\begin{aligned} x_1(o) &= 0.0 \\ x_2(o) &= 0.0 \\ x_3(o) &= 4.0 \end{aligned}$$

<u>a priori</u> distribution:

$m_1 = 0.0$	$p_{11}(o) = 0.0$	$p_{22}(o) = 0.0$
$m_2 = 0.0$	$p_{12}(o) = 0.0$	$p_{23}(o) = 0.0$
$m_3 = 1.0$	$p_{13}(o) = 0.0$	$p_{33}(o) = 4.0$

noise levels:

$$E[w_1(j)w_1(k)] = 1.0 \ \delta_{jk}$$
$$E[w_2(j)w_2(k)] = 0.0$$
$$E[v(j)v(k)] = 0.01 \ \delta_{jk}$$

w_1, w_2 and v are statistically independent.

Discussion: For this data set there is no known input. The natural frequency x_3 is an unknown constant. The noise condition is very severe as indicated by part A of Plate IV. The estimate of x_3 takes about 750 sampling periods to overcome the error in the a priori estimate and settle on a good estimate of x_3. Once this has occurred, the estimates of the other state variables also improve. The total computer time used for the entire simulation was 0.64 minutes.

5.43 Case V

The following data pertains to the simulation results shown in Plate V.

system parameters:

$$a = 0.0$$
$$b = 0.5$$
$$c_o = 0.0$$
$$h = 0.01$$

initial conditions:

$$x_1(o) = 0.0$$
$$x_2(o) = 0.0$$
$$x_3(o) = 3.0$$

a priori distribution:

$m_1 = 0.0$	$p_{11}(o) = 1.0$	$p_{22}(o) = 0.5$
$m_2 = 0.0$	$p_{12}(o) = 0.0$	$p_{23}(o) = 0.0$
$m_3 = 0.0$	$p_{13}(o) = 0.0$	$p_{33}(o) = 4.0$

noise levels:

$$E[w_1(j)w_1(k)] = 0.25 \ \delta_{jk}$$
$$E[w_2(j)w_2(k)] = 0.0$$
$$E[v(j)v(k)] = 1.0 \ \delta_{jk}$$

w_1, w_2 and v are statistically independent.

Discussion: For this data set the natural frequency of the system is again an unknown constant; but this time there is a known input. The noise levels are moderate as indicated by parts B and C of Plate V. Again notice that the estimate of x_3 overcomes the error in the a priori estimate soon after the known input excites a transient. The overall performance of the estimator is quite satisfactory. The behavior of the $\underset{\sim}{P}^*$ matrix is typical. Note that the "variance" p_{33}^* decreases rapidly as the estimate $x_3^*(k|k)$ improves. The total computer time used for the entire simulation was 0.68 minutes.

5.44 Case VI

The following data pertains to the simulation results shown in Plate VI.

system parameters:

$$
\begin{aligned}
b &= 0.5 \\
c_o &= 2.0 \\
h &= 0.01 \\
\text{(system) } a &= 0.1 \\
\text{(estimator) } a &= 0.2
\end{aligned}
$$

initial conditions:

$$
\begin{aligned}
x_1(o) &= 0.5 \\
x_2(o) &= 0.5 \\
x_3(o) &= 7.0
\end{aligned}
$$

a priori distribution:

$$
\begin{aligned}
m_1 &= 0.0 & p_{11}(o) &= 0.5 & p_{22}(o) &= 1.0 \\
m_2 &= 0.0 & p_{12}(o) &= 0.0 & p_{23}(o) &= 0.0 \\
m_3 &= 0.0 & p_{13}(o) &= 0.0 & p_{33}(o) &= 4.0
\end{aligned}
$$

noise levels:

$$
\begin{aligned}
E\left[w_1(j)w_1(k)\right] &= 0.36\ \delta_{jk} \\
\text{(system) } E\left[w_2(j)w_2(k)\right] &= 0.0 \\
\text{(estimator) } E\left[w_2(j)w_2(k)\right] &= 225.0\ \delta_{jk} \\
E\left[v(j)v(k)\right] &= 1.0\ \delta_{jk}
\end{aligned}
$$

w_1, w_2 and v are statistically independent.

<u>Discussion</u>: For this data set the estimator had imperfect information about the statistical behavior of x_3. The actual value was the geometric curve obtained as the output of the following system;

$$x_3(t+h) = x_3(t) - 0.1\, x_3(t)h \qquad\qquad x_3(o) = 7.0$$

The estimator was designed under the assumption that $x_3(t)$ was the output of a linear system of a different time constant excited by white noise.

The noise conditions were moderate to severe as is indicated by parts B and C of Plate VI. Despite the large error in the <u>a priori</u> estimate and the misinformation about the statistical nature of x_3, the estimator quickly overcomes the initial error and gives satisfactory performance, even before the known input u excites a transient in the system. The total computer time used for this simulation was 0.76 minutes.

5.5 Conclusions

In all cases studied the elements of the $\underline{P}^*$ matrix were well behaved. In fact, they were so well behaved that it seems likely that in an application of these techniques, a simplified version of the filter could be obtained by presetting the values of the diagonal elements of the $\underline{P}^*$ matrix to agree with the average behavior of these elements in previously conducted simulation studies. This would be very desirable if an analogue computer were to be used, since it would reduce the number of multipliers needed.

No instability of the estimator was observed.

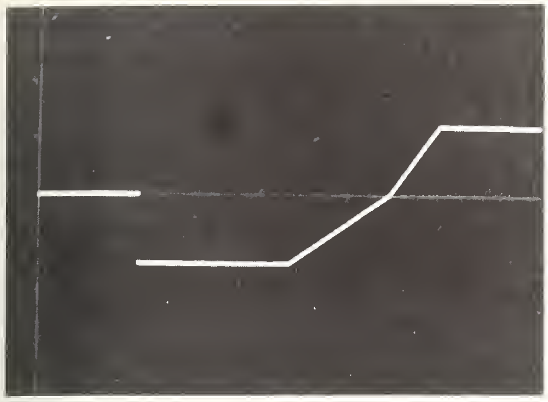
The effect of known inputs is to excite transients which contain information about the system parameters and simplify the estimation problem. The exact nature of the known input does not seem to matter as

long as transients are excited.

The performance of the estimator in Case VI indicates than an estimator designed on the assumption that unknown parameters are the outputs of linear systems excited by white noise, can give good estimates even when this assumption is apparently unjustified. This is very important since information about the statistical behavior of a parameter may be limited. This is also an indication of the power of the model which treats random parameters as the outputs of linear systems excited by white noise. In practice the time constants of these fictitious linear systems would be chosen to reflect the anticipated rapidity of the environmental changes.

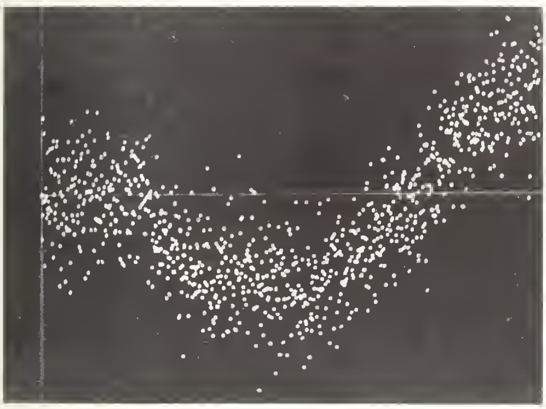
The quality of the estimates depends on the difficulty of the problem; that is, on the presence of known inputs and on the noise levels. The examples presented, for the most part, represent cases of rather severe noise conditions in which satisfactory estimates were repeatedly given. A noisy noise annoys an oyster but apparently, the nonlinear estimator is a pearl of a different sort.

The method of successive linearization converged rapidly to the solution of the nonlinear two-point boundary value problem in the cases investigated. The first approximation $\{\underline{x}^*(k|n)\}$ was excellent in all cases and could be used with confidence as the starting point in any method of successive approximations.

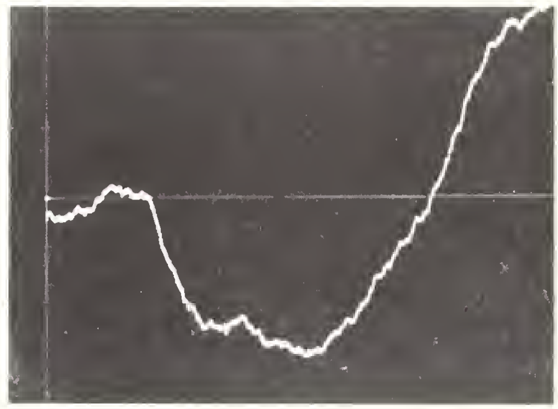


A: Known input $u(k)$
Max: 10.0

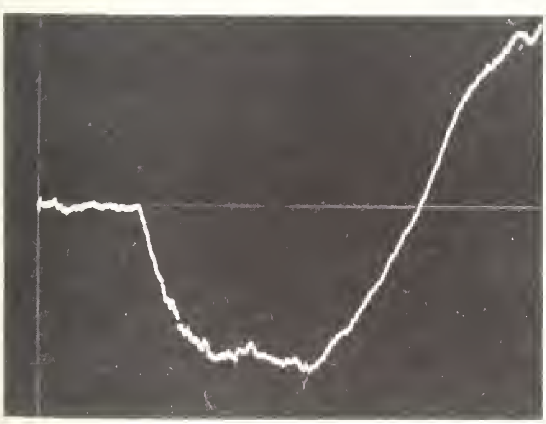
B: Total input $u(k) + w(k)$
Scale of A



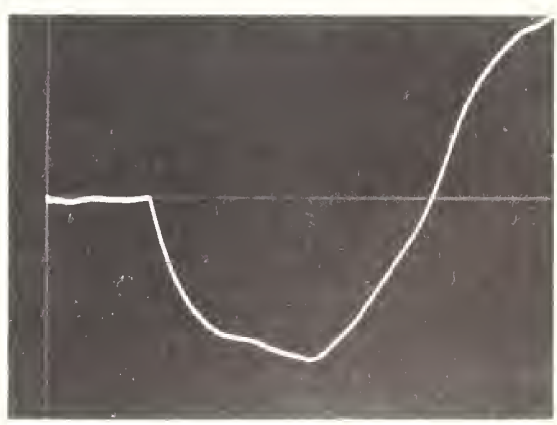
C: Observed output $z(k)$
Max: 16.5



D: State variable $x_1(k)$
Max: 9.5

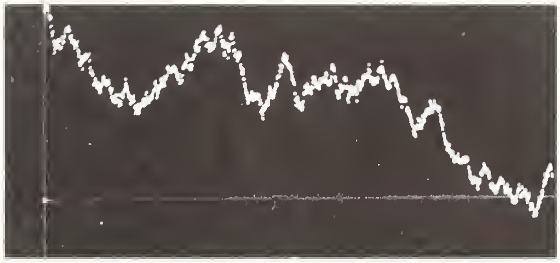


E: Real Time Estimate $x_1^*(k|k)$
Scale of D

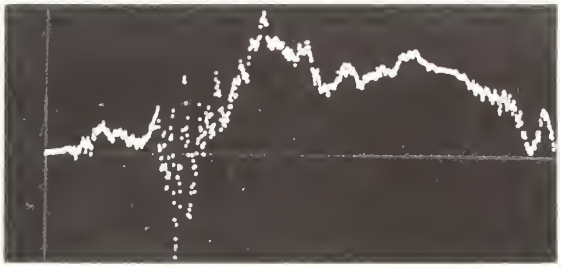


F: Successive Linearization $x_1(k|n)$
$i = 1, 2, 3$ Scale of D

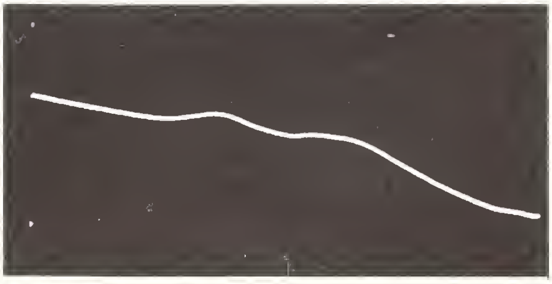
PLATE I



G: State variable $x_2(k)$
Max: 0.51

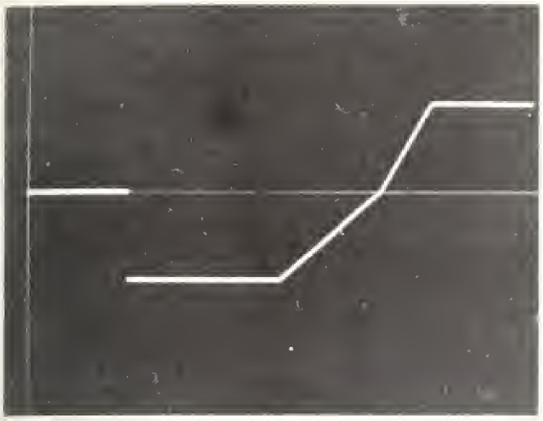


H: Real time estimate $x_2^*(k|k)$
Scale of G



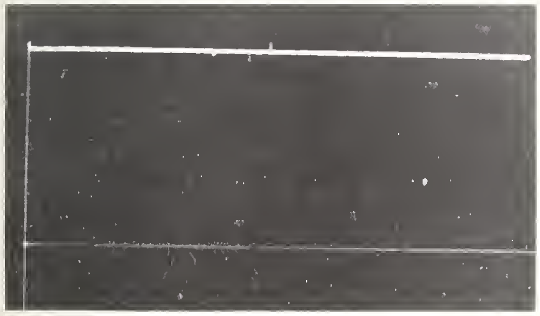
I: Successive linearization $x_{2_i}(k|n)$
$i = 1, 2, 3$ Scale of G

PLATE I (con't.)

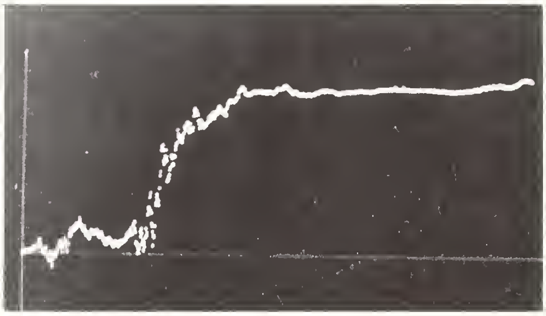


A: Known input $u(k)$
Max: 10.0

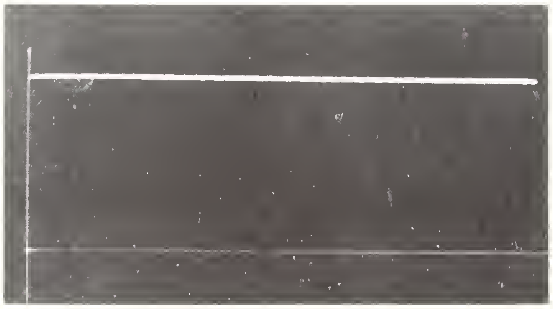
B; Total input $u(k) + w(k)$
Scale of A



C: State variable $x_2(k)$
$x_2(k) = -1.5$



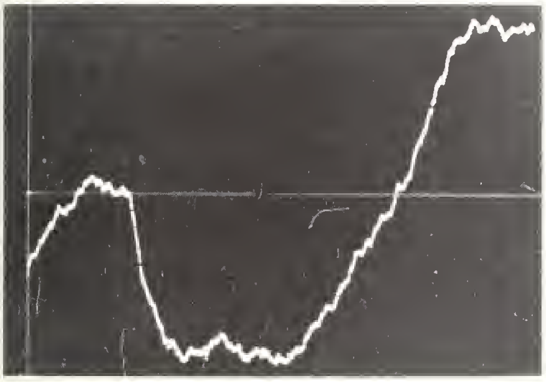
D; Real time estimate $x_2^*(k|k)$
Scale of C



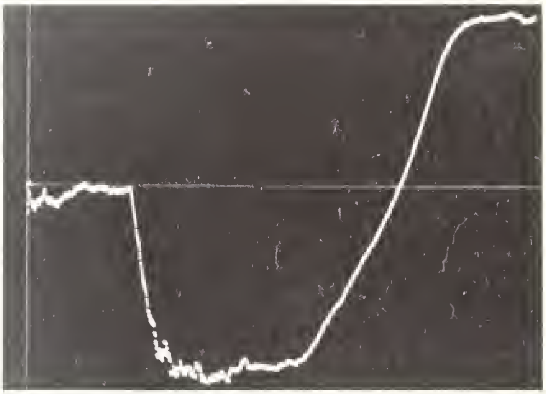
E: Successive Linearization $x_{2_i}(k|n)$
$i = 1, 2, 3$ Scale of C

PLATE II

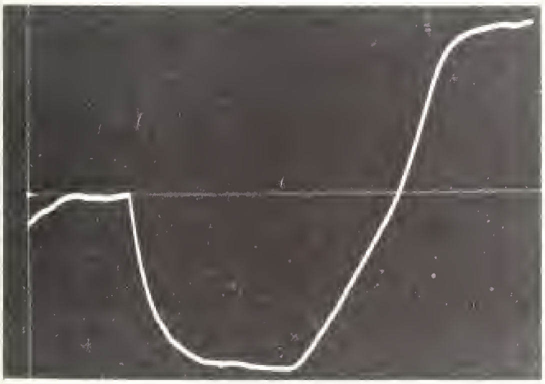
F: Observed output z(k)
Max: 13.4



G: State variable $x_1(k)$
Max: 4.4

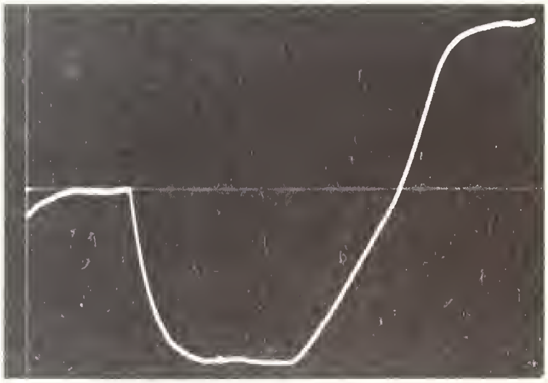


H: Real time estimate $x_1^*(k|k)$
Scale of G



I: Successive linearization
$x_{1_i}(k|n)$

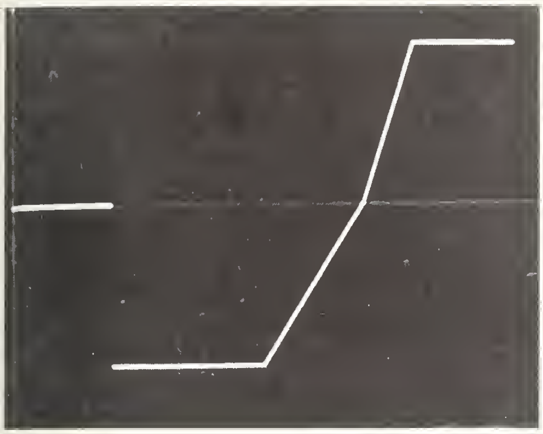
i=1 Scale of G



J: Successive linearization
$x_{1_i}(k|n)$

i = 2, 3 Scale of G

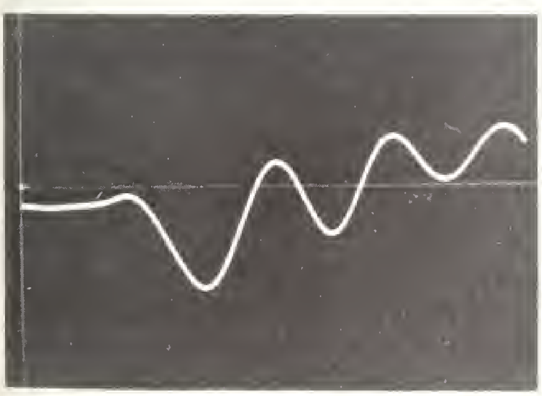
<u>PLATE</u> <u>II</u> (<u>con't</u>.)



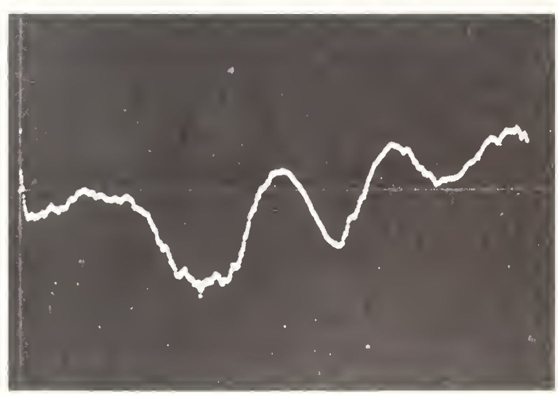
A: Known input u(k)
Max: 5.0

B: Total input u(k) + w(k)
Scale of A

C: Observed output $z(k) = x_1(k) + v(k)$
Max: 4.97

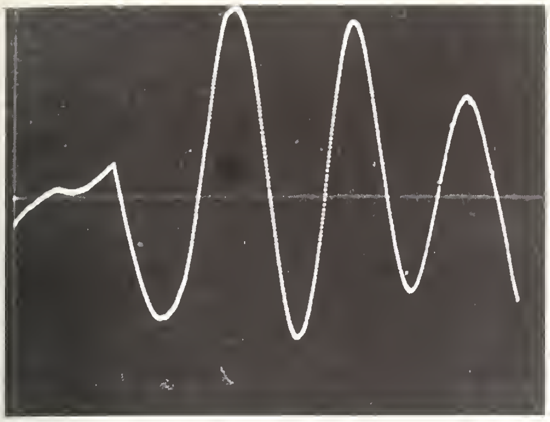


D: State variable $x_1(k)$
Scale of C



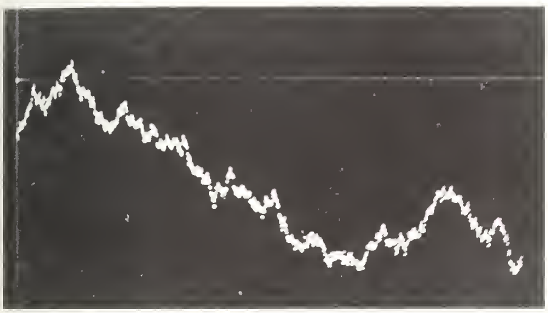
E: Real time estimate $x_1^*(k|k)$
Scale of C

PLATE III

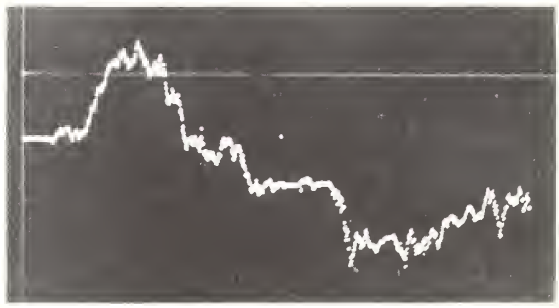


F: State variable $x_2(k)$
Max: 3.5

G: Real time estimate $x_2^*(k|k)$
Scale of F



H: $c(k) = c_o + x_3(k)$
Max: 8.85 , $c_o = 3.0$



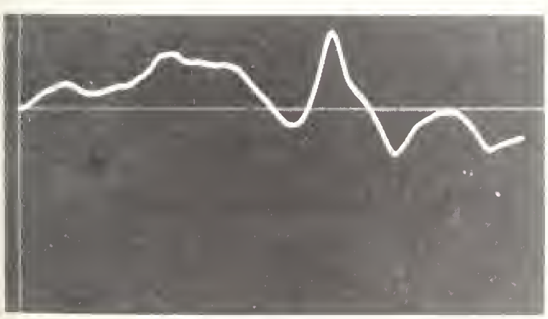
I: Real time estimate
$c^*(k|k) = c_o + x_3^*(k|k)$
Scale of H

J: $P_{11}^*(k)$ Scale of O

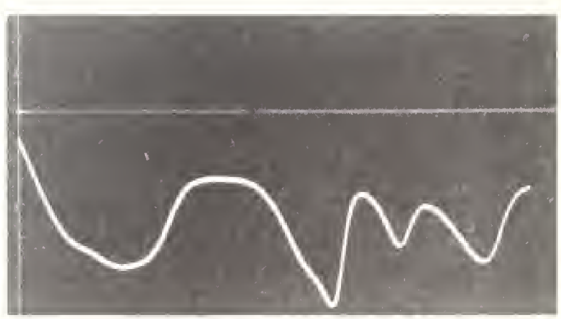
K: $P_{12}^*(k)$ Scale of O

L: $P_{13}^*(k)$ Scale of O

M: $P_{22}^*(k)$ Scale of O

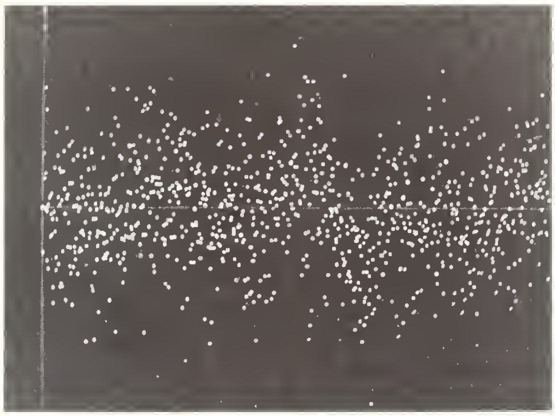


N: $P_{23}^*(k)$ Scale of O

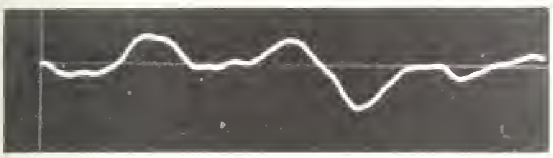


O: $P_{33}^*(k)$ Max: 3.24

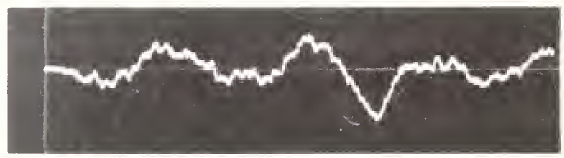
PLATE III (con't.)



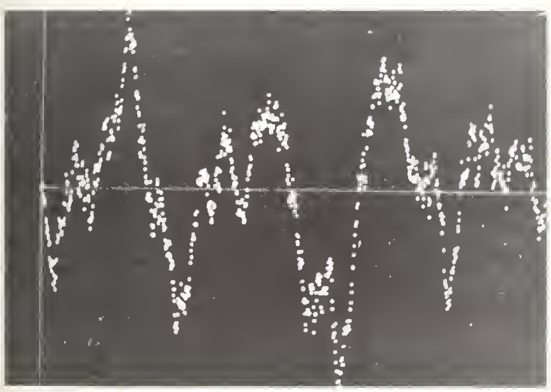
A: Observed output $z(k) = x_1(k) + v(k)$
Max: 0.4



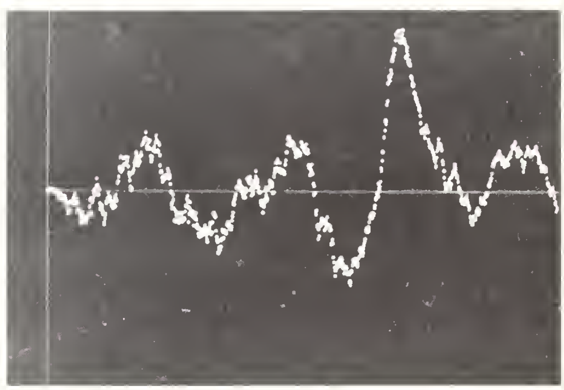
B: State variable $x_1(k)$
Scale of A



C: Real time estimate $x_1^*(k|k)$
Scale of A

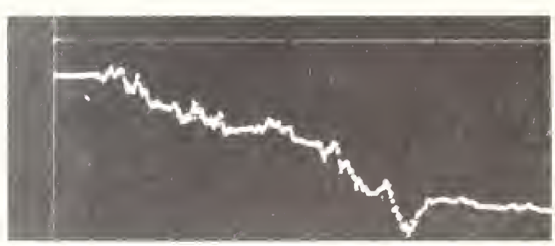


D: State variable $x_2(k)$
Max: 0.2



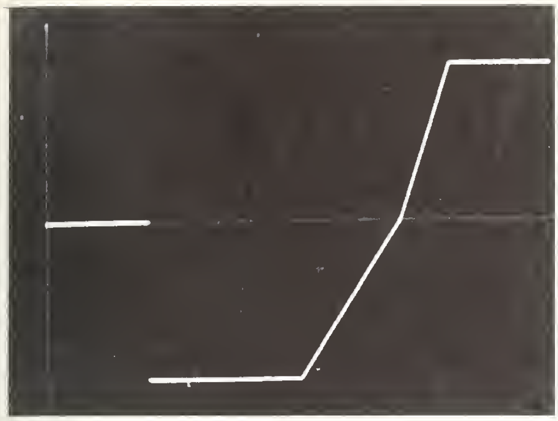
E: Real time estimate $x_2^*(k|k)$
Scale of D

F: State variable $x_3(k)$
Value + 4.0

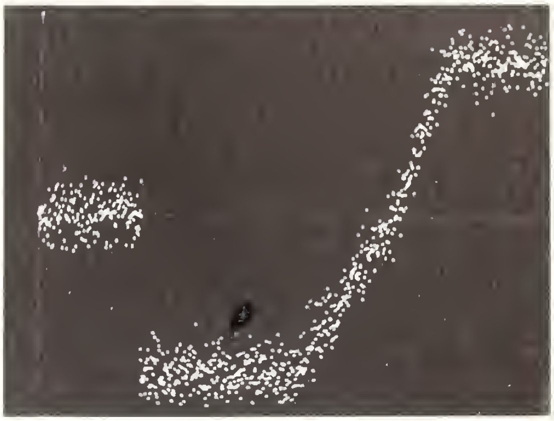


G: Real time estimate $x_3^*(k|k)$
Scale of F

PLATE IV

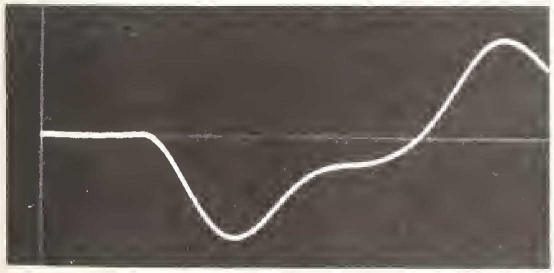


A: Known input u(k)
Max: 5.0

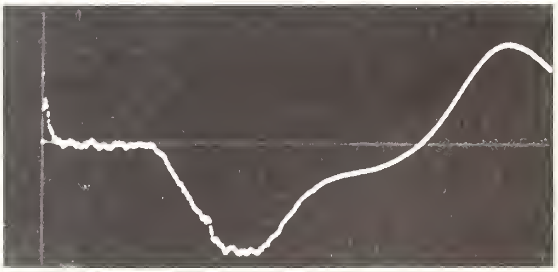


B: Total input u(k) + w(k)
Scale of A

C: Observed output $z(k) = x_1(k) + v(k)$
Max: 5.28

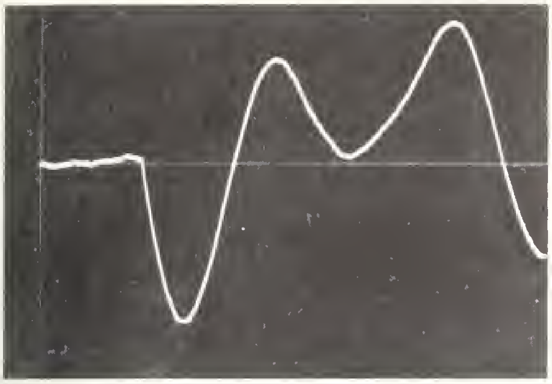


D: State variable $x_1(k)$
Scale of C

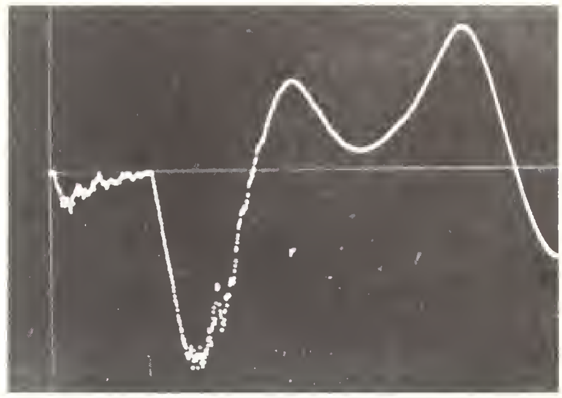


E: Real time estimate $x_1^*(k|k)$
Scale of C

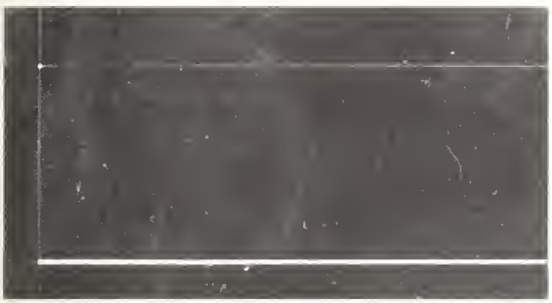
PLATE V



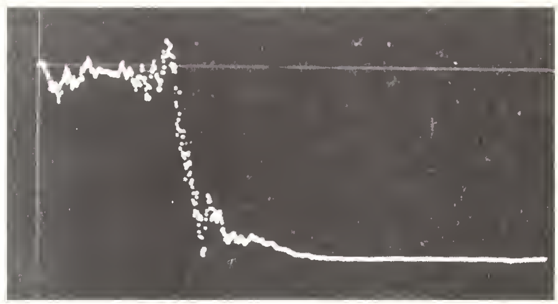
F: State variable $x_2(k)$
Max: 2.44



G: Real time estimate $x_2^*(k|k)$
Scale of F



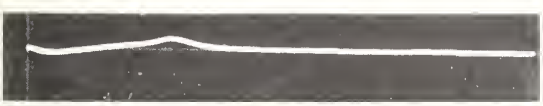
H: State variable $x_3(k)$
Value 3.0



I: Real time estimate $x_3^*(k|k)$
Scale of H

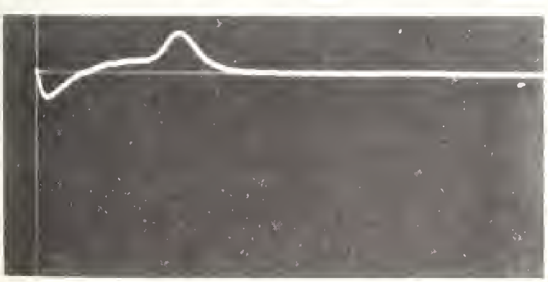
J: $P_{11}^*(k)$ Scale of O

K: $P_{12}^*(k)$ Scale of O

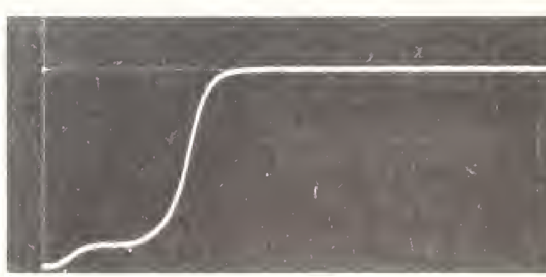


L: $P_{13}^*(k)$ Scale of O

M: $P_{22}^*(k)$ Scale of O



N: $P_{23}^*(k)$ Scale of O

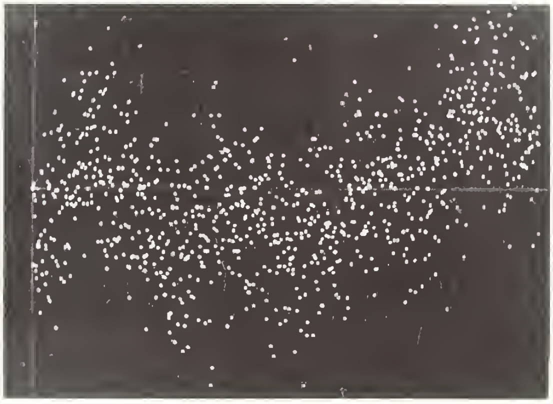


O: $P_{33}^*(k)$ Max: 4.0

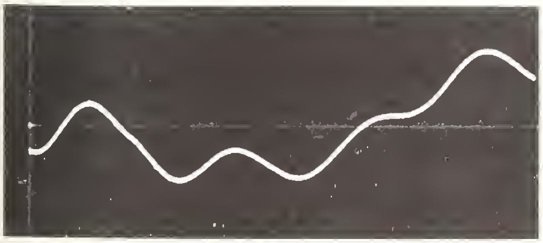
PLATE V (con't.)

A: Known input $u(k)$
Max: 5.0

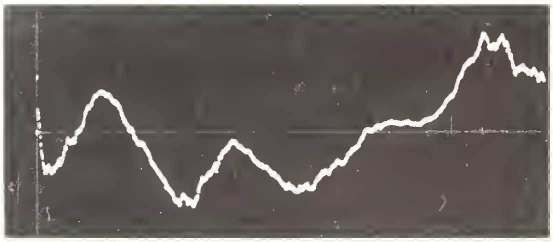
B: Total input $u(k) + w(k)$
Scale of A



C: Observed output $z(k) = x_1(k) + v(k)$

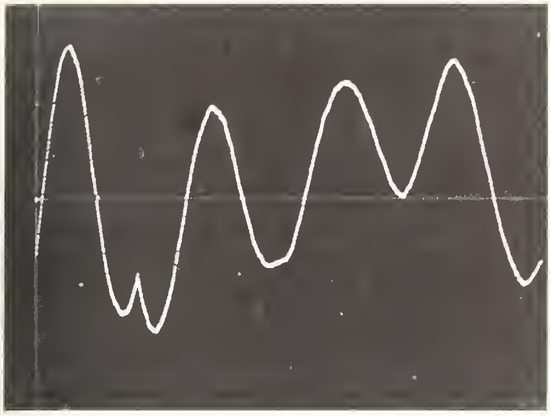


D: State variable $x_1(k)$
Scale of C

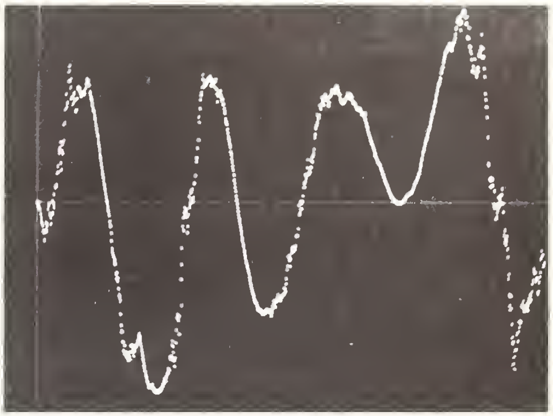


E: Real time estimate $x_1^*(k|k)$
Scale of C

PLATE VI

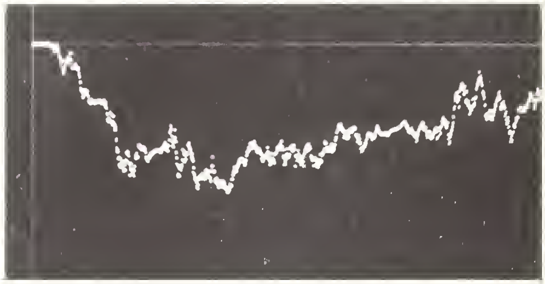


F: State variable $x_2(k)$
Max: 1.4



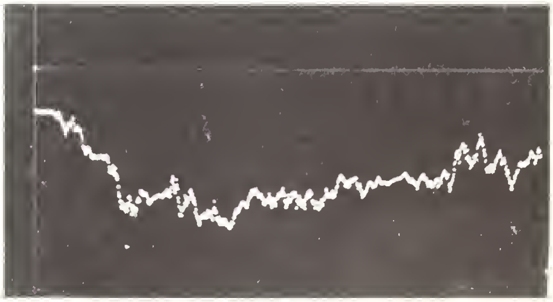
G: Real time estimate $x_2^*(k|k)$
Scale of F

H: State variable $x_3(k)$
Max: 7.0



I: Real time estimate $x_3^*(k|k)$
Scale of H

J: $c(k) = c_o + x_3(k)$
Max: 9.0 , $c_o = 2.0$



K: Real time estimate
$c^*(k|k) = c_o + x_3^*(k|k)$
Scale of J

PLATE VI (con't.)

CHAPTER VI

CONCLUSION

Our research has been devoted to a fundamental problem of considerable importance in modern control theory. The linear problem has been solved (Appendix D) and a scratch has been made in the armor of nonlinearity.

The experimental results of Chapter V are extremely encouraging and indicate that the approximation techniques may have wide applicability. The results show that better performance can be expected when some inputs to the system are known. Such known inputs usually excite transients, the effect of which can be measured at the output. These transients contain information about the state of the system and the values of various parameters. One can further speculate that if information about a particular parameter is not available in the observations, then the performance of the system is relatively insensitive to variations of that parameter and the need for accurate estimation is diminished.

The approximate solution to the nonlinear filtering problem is immediately applicable to adaptive control problems. Note that if measurements can be made of the input and the output of a subsystem containing a particularly troublesome parameter, then the order of the problem can be reduced significantly by treating only that subsystem.

The assumption that various noises are Gaussian is equivalent to making a least-squares estimate when only means and covariances are known.

Like most research programs, this report raises more questions than it answers. It is hoped that the following list will stimulate thought, discussion and future research.

A theoretical question of great practical importance is that of stability. The stability of the linear filter has been shown by Kalman [29] using the direct method of Liapunov [40]. It seems likely that the stability of the nonlinear filter obtained by linearization about the present estimates can be proved under certain restrictions. What are these restrictions?

It would be interesting to know the most general types of systems for which the probability density function (2.7) is unimodal. Convexity of (2.8) is a sufficient condition.

A fast simple numerical procedure for solving the smoothing problem guaranteeing convergence would be of great value.

The areas of observability and controllability are suitable for theoretical research.

If a stationary random process is viewed as the output of a linear system excited by white noise, then the approximation techniques might be used to identify the parameters of this fictitious system. It would be interesting to compare the results of such an approach with the more conventional approach of measuring the correlation function or the spectrum and then approximating the spectrum by a rational polynomial.

It seems likely that a combination of the method of successive linearization and the filtering technique could be used to obtain better estimates in real time. That is, instead of linearizing about the present estimate, one could smooth over the past N observations. This seems to be a real possibility in applications where sampling periods and computation times are of the same order of magnitude.

The development of approximation techniques for use in conjunction with the dynamic programming formulation of the discrete-time problem

would be an important contribution. Such techniques would be immediately applicable to a wide range of dynamic programming problems.

It would be interesting to apply the successive linearization technique to the nonlinear regulator problem discussed in section 3.8.

It would be desirable to simplify the nonlinear filter as much as possible without impairing performance. Presetting the values of the diagonal elements of the $\underline{P}^*$ matrix is one possibility. Another possibility is to use the basic configuration of the filter but to set certain parameters by the method of stochastic approximation [38, 56].

Finally, there is need for further simulation studies applying the methods of this report to more complicated systems.

APPENDIX A

MATRIX IDENTITIES

A1 Quadratic Forms

We begin by summarizing a few facts about matrices so that they will be available for reference.

The length (or Euclidian norm) of a vector $\underline{x}$ is simply

$$\| \underline{x} \| = \sqrt{x_1^2 + \cdots + x_n^2}$$

Clearly, $\| \underline{x} \| > 0$ for $\underline{x} \neq \underline{0}$.

Let $Q(\underline{x})$ be the quadratic form associated with a symmetric matrix $\underline{A}$. That is,

$$Q(\underline{x}) = \underline{x}'\underline{Ax} = \sum_{i,j} a_{ij} x_i x_j \qquad , \; a_{ij} = a_{ji}$$

$\underline{A}$ is said to be positive definite if $Q(\underline{x}) > 0$, for $\underline{x} \neq \underline{0}$. Clearly, $Q(\underline{0}) = 0$. $\underline{A}$ is non-negative definite if $Q(\underline{x}) \geq 0$, for $\underline{x} \neq \underline{0}$. Let $\underline{A}$ and $\underline{B}$ be non-negative definite and let $\underline{C} = \underline{A} + \underline{B}$. Then

$$\underline{x}'\underline{Cx} = \underline{x}'(\underline{A} + \underline{B})\underline{x} = \underline{x}'\underline{Ax} + \underline{x}'\underline{Bx}$$

Thus, $\underline{C}$ is at least non-negative definite. $\underline{C}$ would always be positive definite if either $\underline{A}$ or $\underline{B}$ were positive definite and the other were non-negative definite.

Let D be any matrix. Then $\underline{D}'\underline{D}$ is non-negative definite, since $\underline{x}'\underline{D}'\underline{Dx} = \| \underline{Dx} \|^2 \geq 0$. If D is non-singular, then $\underline{D}'\underline{D}$ is positive definite since in this case $\underline{Dx} = \underline{0}$ implies $\underline{x} = \underline{0}$.

Conversely, any positive definite matrix $\underline{A}$ may be factored into a product of the form $\underline{D}'\underline{D}$ where $\underline{D}$ is non-singular. To see this we recall that the characteristic numbers of a positive definite matrix are positive and that any symmetric matrix may be reduced by an orhtogonal transformation to a diagonal matrix having its characteristic numbers on the diagonal. Let $\underline{A}$ be positive definite and let $\underline{T}$ be the appropriate orthogonal transformation ($\underline{T}' = \underline{T}^{-1}$). Then

$$\underline{T}'\underline{AT} = \underline{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_n) \qquad , \lambda_i > 0$$

A suitable $\underline{D}'$ is $\underline{T}' \,\mathrm{diag}(\sqrt{\lambda}_1, \dots, \sqrt{\lambda}_n)$. If $\underline{A}$ were non-negative definite $\underline{D}$ would be singular, since some of the terms on the diagonal of $\underline{\Lambda}$ would be zero. If $\underline{A}$ were positive definite, then $\underline{A}^{-1}$ could be written in the following form;

$$\underline{A}^{-1} = \underline{T}\,\mathrm{diag}(1/\lambda_1, \dots, 1/\lambda_n)\underline{T}'$$

which shows that $\underline{A}^{-1}$ is positive definite.

Finally, if $\underline{A}$ is non-negative, then $\underline{H}'\underline{AH}$ is at least non-negative definite, since $\underline{x}'\underline{H}'\underline{AHx} = \|\underline{DHx}\|^2 \geq 0$.

A2 A Visual Approach

In the opinion of the author, the use of block diagrams or flow graphs enables one to produce matrix identities far faster and more simply than by straight algebraic techniques. We will view a matrix as a linear operator. For example, the equation $\underline{y} = \underline{Ax}$ we interpret as a system $\underline{A}$ operating on an input $\underline{x}$ to produce a unique output $\underline{y}$. Similarly, the equation $\underline{y} = \underline{ABx}$ is given the interpretation that first $\underline{B}$ operates on $\underline{x}$ and then $\underline{A}$ operates on this result to produce $\underline{y}$. See Fig. A-1.

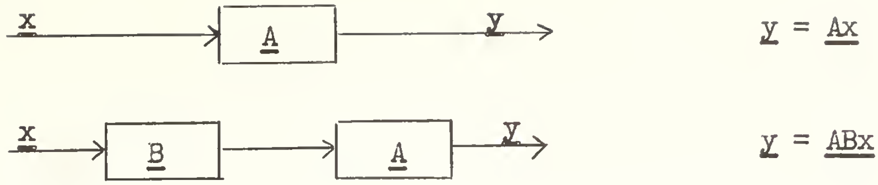


Fig. A-1

Some basic axioms of matrix theory and the associated block diagram manipulations are given in Fig. A-2.

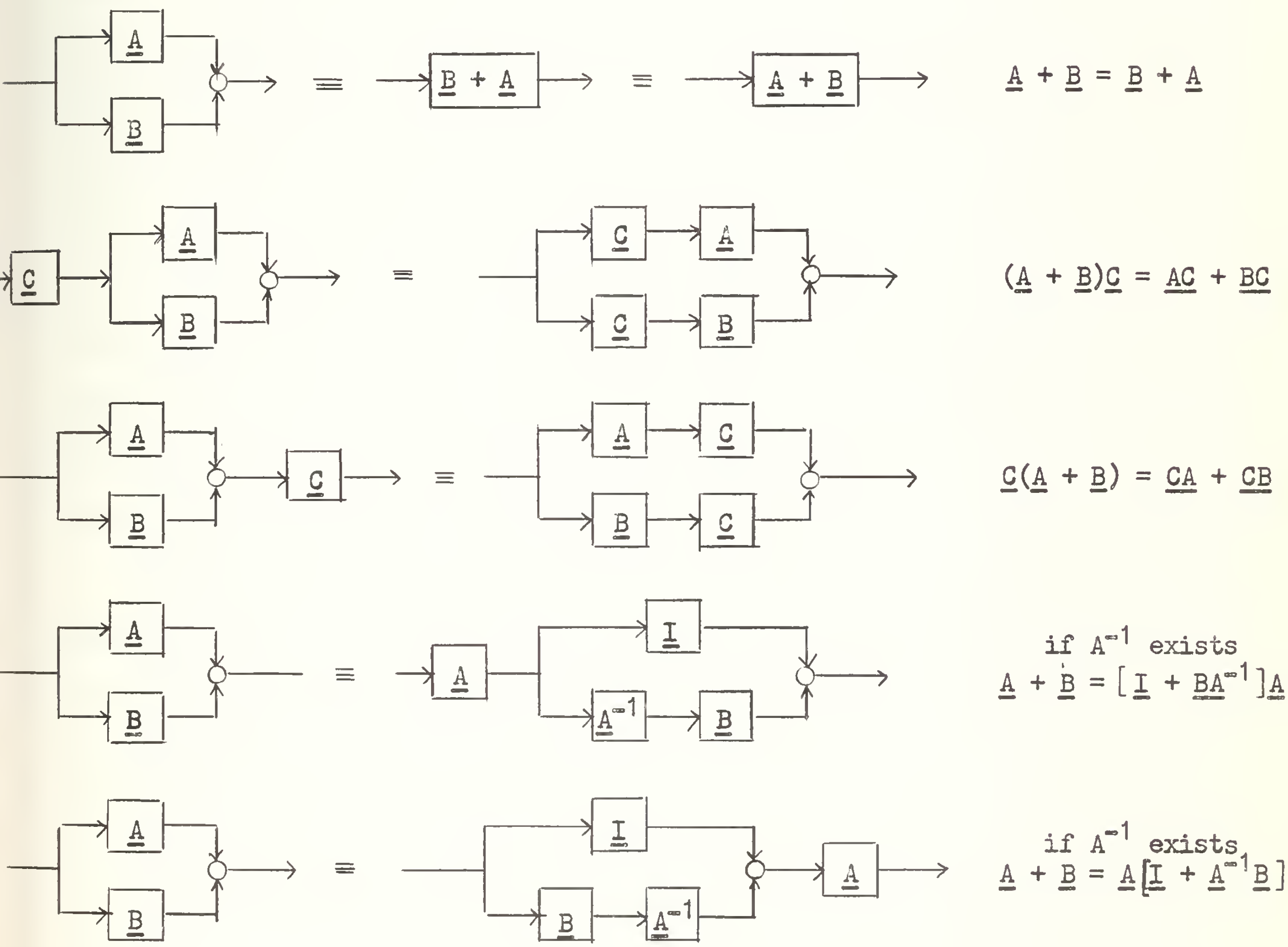


Fig. A-2

The question of placing a matrix inside a feedback loop is slightly more complex but is of extreme importance. Consider the "system" of A-3.

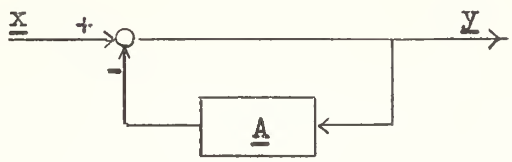


Fig. A-3

This "system" implies the equation $\underline{y} = \underline{x} - \underline{A}\underline{y}$ or $[\underline{I} + \underline{A}]\underline{y} = \underline{x}$. Note that the output $\underline{y}$ is specified uniquely by $\underline{x}$ only if $[\underline{I} + \underline{A}]^{-1}$ exists. Thus, the diagram of Fig. A-3 is valid only if $[\underline{I} + \underline{A}]^{-1}$ exists, since only then is the output uniquely specified by the input and in our basic concept of a system we require this.

In order for a block diagram to represent a valid system the output of each individual element of the block diagram must be uniquely specified for each possible input. The block diagram manipulations of Fig. A-2 may then be performed without affecting this uniqueness to obtain equivalent systems.

Some examples will now be given to show how the ability to picture a matrix operation in the form of a block diagram is of great assistance. Important relations are obtained in Example 3.

Example 1. Show that

$$[\underline{I} + \underline{AB}]^{-1} = [\underline{A}^{-1} + \underline{B}]^{-1}\underline{A}^{-1} \qquad \text{if } \underline{A}^{-1} \text{ exists}$$

and

$$[\underline{I} + \underline{AB}]^{-1} = \underline{B}^{-1}[\underline{B}^{-1} + \underline{A}]^{-1} \qquad \text{if } \underline{B}^{-1} \text{ exists}$$

We use the visual aid of the block diagram of Fig. A-4, although a straightforward algebraic proof is simple for this simple problem.

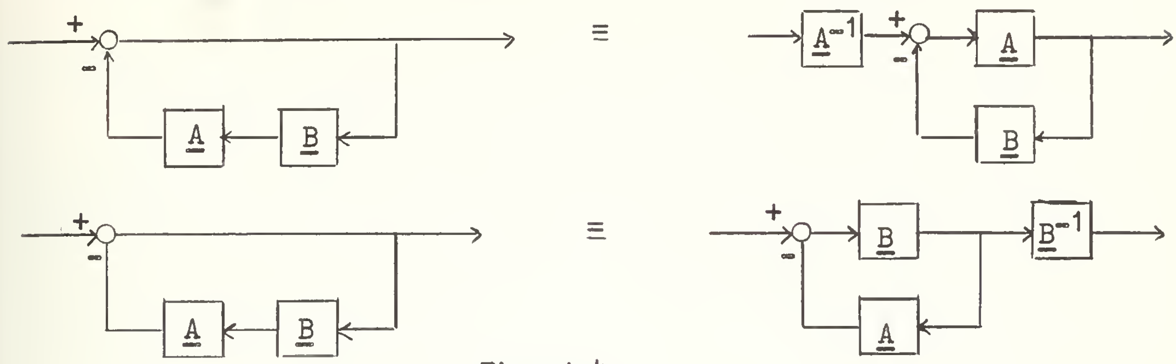


Fig. A-4

Note that in moving blocks around the loop it is clear when the additional assumption must be made as to the existence of some inverse.

<u>Example 2.</u> Consider the question, "Does the existence of $[\underline{I} + \underline{AB}]^{-1}$ imply the existence of $[\underline{I} + \underline{BA}]^{-1}$?" If so, determine the latter in terms of the former. We note that from Fig. A-5 $[\underline{I} + \underline{AB}]^{-1}\underline{A} = \underline{A}\,[\underline{I} + \underline{BA}]^{-1}$, so the answer to the question is yes.

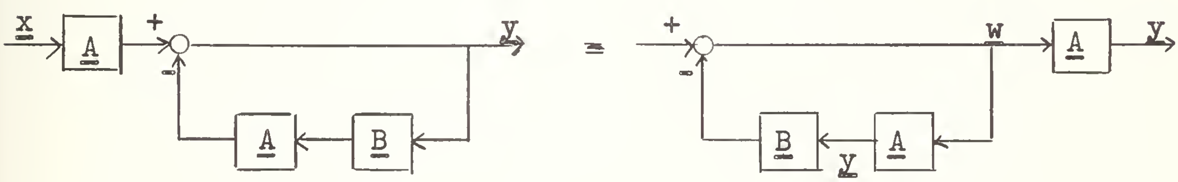


Fig. A-5

To obtain an expression for the inverse we introduce the quantity $\underline{w} = \underline{x} - \underline{By}$. Then, by the definition of $\underline{y}$ in the diagram on the left side of Fig. A-5,

$$\underline{w} = [\underline{I} - \underline{B}\,[\underline{I} + \underline{AB}]^{-1}\underline{A}]\,\underline{x}$$

but, by the diagram on the right side of Fig. A-5,

$$\underline{w} = [\underline{I} + \underline{BA}]^{-1}\underline{x}$$

for all $\underline{x}$. Thus

$$[\underline{I} + \underline{BA}]^{-1} = \underline{I} - \underline{B}\,[\underline{I} + \underline{AB}]^{-1}\,\underline{A} \tag{A.1}$$

which may be verified by direct multiplication.

Example 3. As a third example we shall develop two matrix identities of practical importance. Suppose $\underline{R}^{-1}$ exists and consider the expression $\underline{PH}'\,[\,\underline{HPH}' + \underline{R}]^{-1}$. In Fig. A-6 we simply back things around the loop to obtain the first desired result by inspection.

$$\underline{PH}'\,[\,\underline{HPH}' + \underline{R}]^{-1} = [\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H}]^{-1}\;\underline{PH}'\underline{R}^{-1} \tag{A.2}$$

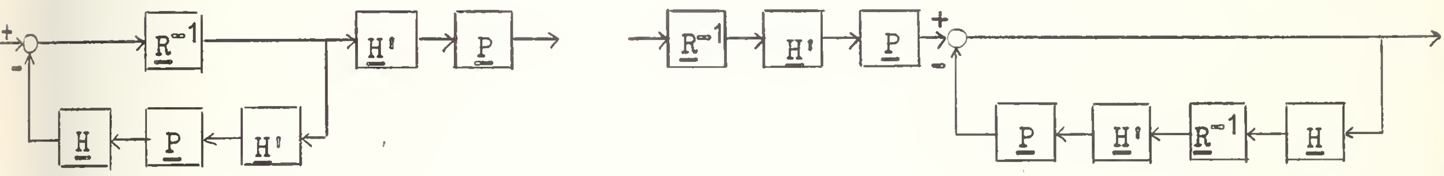


Fig. A-6

Using (A.2) we note that

$$\underline{I} - \underline{PH}'\,[\,\underline{HPH}' + \underline{R}]^{-1}\,\underline{H} = \underline{I} - [\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H}]^{-1}\;\underline{PH}'\underline{R}^{-1}\underline{H}$$

Rewriting $\underline{I}$ as a product, this relation is equal to

$$[\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H}]^{-1}\;[\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H}] - [\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H}]^{-1}\;\underline{PH}'\underline{R}^{-1}\underline{H}$$

Canceling terms, we obtain the identity

$$\underline{I} - \underline{PH}'[\,\underline{HPH}' + \underline{R}]^{-1}\,\underline{H} = [\,\underline{I} + \underline{PH}'\underline{R}^{-1}\underline{H}]^{-1} \tag{A.3}$$

This result may be verified by direct multiplication. Note that (A.3) could also have been obtained by substituting $\underline{PH}'$ for $\underline{B}$ and $\underline{R}^{-1}\underline{H}$ for $\underline{A}$

in (A.1) and then bringing $\underline{R}^{-1}$ inside the inverse.

The easy way in which the results (A.2) and (A.3) were obtained using the graphic aid of the block diagram is an indication of the power of this approach both for obtaining identities and proving the existence of inverses.

APPENDIX B

THE GENERALIZED INVERSE OF A MATRIX

B1 Introduction

The concept of the inverse of a matrix was first generalized by Moore [47, 48] to include rectangular and singular square matrices. Later, Penrose [50, 51] independently introduced the same basic concept. Further discussion of the generalized inverse was given by Greville [21, 22].

Penrose proves that for any matrix $\underline{A}$ there is a unique matrix $\underline{A}^{\#}$ satisfying the following four relations;

$$\underline{A}\underline{A}^{\#}\underline{A} = \underline{A} \tag{B.1}$$

$$(\underline{A}\underline{A}^{\#})' = \underline{A}\underline{A}^{\#} \tag{B.2}$$

$$\underline{A}^{\#}\underline{A}\underline{A}^{\#} = \underline{A}^{\#} \tag{B.3}$$

$$(\underline{A}^{\#}\underline{A})' = \underline{A}^{\#}\underline{A} \tag{B.4}$$

and proceeds to discuss the properties of the generalized inverse.

Kalman [29] defines a pseudo-inverse of a matrix $\underline{A}$ to be any matrix $\underline{A}^{\dagger}$ satisfying (B.1) and shows how such a pseudo-inverse may be used in linear filtering and prediction problems.

This appendix is intended to serve as an introduction to the notions of generalized inverses and pseudo-inverses. We discuss their use in connection with the solution of linear algebraic equations. A geometrical interpretation of the generalized inverse in terms of orthogonal projections is presented. For simplicity, the discussion is

limited to real matrices.

B2 The Equation $\underline{Ax} = \underline{y}$

Consider the set of linear algebraic equations;

$$a_{11}x_1 + \cdots + a_{1n}x_n = y_1$$

$$\vdots$$

$$a_{m1}x_1 + \cdots + a_{mn}x_n = y_m \tag{B.5}$$

or, equivalently, the matrix equation $\underline{Ax} = \underline{y}$, where $\underline{A}$ is a known m by n matrix, $\underline{y}$ is a known m-vector, and $\underline{x}$ is an n-vector to be determined.

If there is no vector $\underline{x}$ which satisfies this relation, the set (B.5) is said to be inconsistent. Otherwise, the set is consistent and there may be a unique solution or an infinite number of solutions.

Suppose a solution exists and we decide to choose the smallest solution; that is, the solution for which $\|\underline{x}\|$ or, equivalently, $\frac{1}{2}\|\underline{x}\|^2$, is minimum. If we introduce a vector of Lagrange multipliers and treat (B.5) as a constraint, we may minimize

$$\frac{1}{2}\,\underline{x}'\underline{x} + \underline{\lambda}'\,(\underline{y} - \underline{Ax})$$

Setting the partial derivatives of this expression with respect to $\underline{x}$ and $\underline{\lambda}$ equal to zero yields the following two equations;

$$\underline{x} = \underline{A}'\underline{\lambda} \tag{B.6}$$

$$\underline{y} = \underline{Ax} \tag{B.7}$$

Combining these equations, we obtain

$$\underline{y} = \underline{AA}'\,\underline{\lambda} \tag{B.8}$$

If $\underline{AA}'$ is non-singular we may solve this equation for $\underline{\lambda}$ and combine with (B.6) to obtain the desired solution;

$$\underline{x} = \underline{A}'(\underline{AA}')^{-1}\underline{y} \tag{B.9}$$

One can easily verify that if $\underline{A}'(\underline{AA}')^{-1}$ exists, then it satisfies (B.1) through (B.4) and hence, is the generalized inverse.

Suppose, on the other hand, that the set (B.5) is inconsistent. Then one might wish to make a least squares approximation by minimizing $\|\underline{Ax} - \underline{y}\|^2$. If we set the derivative of this expression equal to zero, we obtain

$$\underline{A}'\underline{Ax} = \underline{A}'\underline{y} \tag{B.10}$$

If $\underline{A'A}$ is non-singular we may solve this expression for $\underline{x}$ to obtain the desired result.

$$\underline{x} = (\underline{A}'\underline{A})^{-1}\underline{A}'\underline{y} \tag{B.11}$$

If $(\underline{A}'\underline{A})^{-1}\underline{A}'$ esists, it satisfies (B.1) through (B.4) and hence, is the generalized inverse.

B3 The Pseudo-Inverse

A pseudo-inverse of a matrix $\underline{A}$ is defined [29] as any matrix $\underline{A}^{\dagger}$ satisfying the relation

$$\underline{AA}^{\dagger}\underline{A} = \underline{A} \tag{B.12}$$

It follows from (B.12) that $(\underline{A}')^{\dagger} = (\underline{A}^{\dagger})'$ and that if $\underline{A}^{-1}$ exists, it is equal to $\underline{A}^{\dagger}$.

The following procedure, similar to that given in [29] , may be

used to obtain a pseudo-inverse for any matrix. We first consider the case in which $\underline{A}$ is square. It is well known [17] that any square matrix may be reduced to a diagonal canonical form by a similarity transformation,

$$\underline{PAQ} = \underline{E}$$

where $\underline{P}$ and $\underline{Q}$ are non-singular and $\underline{E}$ is a diagonal matrix having only zeros and ones on the diagonal. Then,

$$\underline{PAQPAQ} = \underline{E}^2 = \underline{E}$$

and

$$\underline{AQPA} = \underline{P}^{-1}\underline{EQ}^{-1} = \underline{A}$$

Hence, $\underline{QP}$ satisfies (B.12) and is a pseudo-inverse of $\underline{A}$. Since $\underline{Q}^{-1}$ and $\underline{P}^{-1}$ exist, a non-singular pseudo-inverse may always be found for a square matrix.

To obtain a pseudo-inverse of a non-square matrix we make use of the fact that either of the following relations imply that $\underline{B}$ is a pseudo-inverse of $\underline{A}$;

$$(\underline{ABA} - \underline{A})(\underline{ABA} - \underline{A})' = \underline{0} \tag{B.13}$$

$$(\underline{ABA} - \underline{A})'(\underline{ABA} - \underline{A}) = \underline{0} \tag{B.14}$$

Direct substitution into (B.13) and (B.14) respectively will verify that the following expressions are pseudo-inverses of $\underline{A}$;

$$\underline{A}'(\underline{AA}')^{\dagger} \tag{B.15}$$

$$(\underline{A}'\underline{A})^{\dagger}\underline{A}' \tag{B.16}$$

The pseudo-inverse in (B.15) and (B.16) may be computed by the first method since ($\underline{A}'\underline{A}$) and ($\underline{AA}'$) are square.

Since (B.15) and (B.16) are, in general, not the same, we see that the pseudo-inverse is not, in general, unique. In fact, if $\underline{B}$ is a pseudo-inverse of $\underline{A}$, then so are

$$\underline{B} + (\underline{I} - \underline{BA})\underline{M} \qquad \text{and} \qquad \underline{B} + \underline{M}(\underline{I} - \underline{AB})$$

where $\underline{M}$ is an arbitrary matrix.

The pseudo-inverse may be used in connection with the solution of equations such as (B.5) because it has the property that any pseudo-inverse of $\underline{A}$ may be used to test for the existence of a solution and any pseudo-inverse of $\underline{A}$ may be used to obtain a solution when one exists. This property is summarized in the following theorem which is similar to one proved by Penrose for the generalized inverse.

Theorem: The equation $\underline{y} = \underline{Ax}$, where $\underline{y}$ and $\underline{A}$ are known, has a solution if and only if, for any $\underline{A}^\dagger$ satisfying (B.12), the following relation is satisfied;

$$\underline{y} = \underline{AA}^\dagger\underline{y} \tag{B.17}$$

If (B.17) is satisfied for some pseudo-inverse $\underline{A}^\dagger$, then it is satisfied for every pseudo-inverse of $\underline{A}$ and $\underline{x} = \underline{A}^\dagger\underline{y}$ is a solution of (B.5) for every pseudo-inverse of $\underline{A}$.

Proof: i., (if). Let $\underline{A}_o^\dagger$ be any pseudo-inverse for which (B.17) holds. Then $\underline{A}_o^\dagger\,\underline{y}$ is a solution of (B.5).

ii., (only if). Suppose $\underline{x}_1$ is a solution of (B.5). Then,

$$\underline{y} = \underline{Ax}_1 = \underline{AA}^\dagger\underline{Ax}_1 = \underline{AA}^\dagger\underline{y}$$

for every pseudo-inverse of $\underline{A}$.

iii., Let $\underline{A}_o^\dagger$ be any pseudo-inverse for which (B.17) is satisfied and let $\underline{A}_n^\dagger$ be any pseudo-inverse for which (B.17) is not satisfied. Then,

$$\underline{y} \neq \underline{AA}_n^\dagger \underline{y} = \underline{AA}_n^\dagger \underline{AA}_o^\dagger \underline{y} = \underline{AA}_o^\dagger \underline{y}$$

which is a contradiction. Hence, if (B.17) holds for some pseudo- inverse, then it holds for every pseudo-inverse and $\underline{A}^\dagger \underline{y}$ is a solution of (B.5) for every pseudo-inverse of $\underline{A}$.

Let us return to the problem of finding the solution of (B.5) having the smallest norm. We are faced with the problem of determining a Lagrange multiplier $\underline{\lambda}$ which satisfies (B.8). Such a Lagrange multiplier will exist by virtue of the theorem, if and only if,

$$\underline{y} = \underline{AA}'(\underline{AA}')^\dagger \underline{y} \tag{B.18}$$

Noting that $\underline{A}'(\underline{AA}')^\dagger$ is a pseudo-inverse of $\underline{A}$, it follows from the theorem that (B.18) will be satisfied and (B.8) will always have a solution if (B.5) has a solution, which was the original hypothesis. Hence, we may use any $(\underline{AA}')^\dagger$ to obtain

$$\underline{\lambda} = (\underline{AA}')^\dagger \underline{y}$$

and finally, using (B.6), the smallest solution of (B.5) is given by

$$\underline{x} = \underline{A}'(\underline{AA}')^\dagger \underline{y} \tag{B.19}$$

This result differs from (B.9) in that we did not have to assume the existence of $(\underline{AA}')^{-1}$. It is interesting to observe that $\underline{A}'(\underline{AA}')^\dagger$ satisfies (B.3) and (B.4) as well as (B.1).

If we return to the problem of finding the least squares approx-

imation, we are faced with finding an $\underline{x}$ which satisfies (B.10). Such an $\underline{x}$ will always exist and we may use any $(\underline{A}'\underline{A})^{\dagger}$ to obtain

$$\underline{x} = (\underline{A}'\underline{A})^{\dagger}\underline{A}'\underline{y} \tag{B.20}$$

which is similar to (B.10). We observe that $(\underline{A}'\underline{A})^{\dagger}\underline{A}'$ satisfies (B.2) and (B.3) as well as (B.1).

B4 The Generalized Inverse

The generalized inverse of a matrix $\underline{A}$ is the unique matrix satisfying (B.1) through (B.4). In the previous section we observed that $\underline{A}'(\underline{AA}')^{\dagger}$ and $(\underline{A}'\underline{A})^{\dagger}\underline{A}'$ each satisfied three out of the four of these relations. Hence, it is not surprising that the generalized inverse may be written in forms resembling these expressions.

We first consider the case in which $\underline{A}$ is symmetric. Then $\underline{A}$ may be reduced by an orthogonal transformation to a diagonal matrix having the characteristic roots of $\underline{A}$ on the diagonal.

$$\underline{T}'\underline{AT} = \underline{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

We form $\underline{\Lambda}^{\#}$ from $\underline{\Lambda}$ by replacing λ_i by $1/\lambda_i$, for $\lambda_i \neq 0$, and leaving unchanged the terms of $\underline{\Lambda}$ which are equal to zero. Then,

$$\underline{A}^{\#} = \underline{T}\ \underline{\Lambda}^{\#}\underline{T}' \tag{B.21}$$

It is simple to verify that $\underline{A}^{\#}$ given by (B.21) satisfies (B.1) through (B.4). Since $\underline{AA}'$ and $\underline{A}'\underline{A}$ are symmetric, we may find $(\underline{AA}')^{\#}$ and $(\underline{A}'\underline{A})^{\#}$ in this manner. One can easily verify that the following two expressions satisfy (B.1) through (B.4) and hence, are equivalent expressions for the generalized inverse of $\underline{A}$;

$$\underline{A}'(\underline{AA}')^{\#} \tag{B.22}$$

$$(\underline{A}'\underline{A})^{\#}\underline{A}' \tag{B.23}$$

It is clear that the generalized inverse is a particular pseudo-inverse and hence, may be used to obtain solutions to linear equations when solutions exist. The generalized inverse, in addition, has the two important properties of always giving the smallest solution when a solution exists and giving the best least squares approximation when no solution exists. These properties are summarized in the following theorem given by Penrose [51] which we state without proof.

Theorem: Let $\| \underline{A} \|$ = trace $\underline{A}'\underline{A}$. $\underline{X}_o$ is said to be the best approximate solution to the matrix equation $\underline{AX} = \underline{Y}$, if, for all $\underline{X}$, either

$$\| \underline{AX} - \underline{Y} \| > \| \underline{AX}_o - \underline{Y} \|$$

or $\| \underline{AX} - \underline{Y} \| = \| \underline{AX}_o - Y \|$ and $\| \underline{X} \| \geq \| \underline{X}_o \|$

Then $\underline{A}^{\#}\underline{Y}$ is the unique best approximate solution to the equation $\underline{AX} = \underline{Y}$.

° ° °

An alternate expression for the generalized inverse may be obtained in the following manner. Reconsider the problem of finding a best least squares approximation to the solution of the equation $\underline{Ax} = \underline{y}$. We found previously that any $\underline{x}$ satisfying $\underline{A}'\underline{Ax} = \underline{A}'\underline{y}$ would do, but we shall now seek the smallest $\underline{x}$ satisfying this relation. We minimize the expression

$$\tfrac{1}{2} \| \underline{x} \|^2 + \underline{\lambda}'(\underline{A}'\underline{Ax} - \underline{A}'\underline{y}) \tag{B.24}$$

Setting the partial derivatives of this expression, with respect to $\underline{x}$ and $\underline{\lambda}$, equal to zero we obtain the following two equations;

$$\underline{A}'\underline{A}\underline{x} = \underline{A}'\underline{y} \tag{B.10}$$

$$\underline{x} = \underline{A}'\underline{A}\ \underline{\lambda} \tag{B.25}$$

These may be combined to obtain

$$\underline{A}'\underline{A}\underline{A}'\underline{A}\ \underline{\lambda} = \underline{A}'\underline{y}$$

Using a pseudo-inverse to solve for $\underline{\lambda}$, we obtain

$$\underline{\lambda} = (\underline{A}'\underline{A}\underline{A}'\underline{A})^{\dagger}\underline{A}'\underline{y} \tag{B.26}$$

Combining (B.25) and (B.26) yields

$$\underline{x} = \underline{A}'\underline{A}(\underline{A}'\underline{A}\underline{A}'\underline{A})^{\dagger}\underline{A}'\underline{y} \tag{B.27}$$

The matrix $\underline{A}'\underline{A}(\underline{A}'\underline{A}\underline{A}'\underline{A})^{\dagger}\underline{A}'$ may be shown to satisfy (B.1) through (B.4) and hence, is an expression for $\underline{A}^{\#}$. In verifying this fact it is helpful to note that $\underline{A}'\underline{A}(\underline{A}'\underline{A}\underline{A}'\underline{A})^{\dagger}$ is a pseudo-inverse of $\underline{A}'\underline{A}$ by (B.15). We omit the details. Then,

$$\underline{A}^{\#} = \underline{A}'\underline{A}(\underline{A}'\underline{A}\underline{A}'\underline{A})^{\dagger}\underline{A}' \tag{B.28}$$

There is also a related expression for $\underline{A}^{\#}$ given by

$$\underline{A}^{\#} = \underline{A}'(\underline{A}\underline{A}'\underline{A}\underline{A}')^{\dagger}\underline{A}\underline{A}' \tag{B.29}$$

Suggestions for computing the generalized inverse are given in [6, 22, 51].

B5 A Geometrical Interpretation of the Generalized Inverse

In order to give the generalized inverse a simple interpretation in terms of orthogonal projections we will need some basic concepts about

linear transformations and vector spaces such as may be found in [2].

We will view the m by n matrix $\underline{A}$ as a set of column vectors $\underline{a}_1,\ldots,\underline{a}_n$. If the equation $\underline{y} = \underline{Ax}$ is written in the following form, emphasizing this point of view,

$$\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \underline{a}_1 & \underline{a}_2 & \cdots & \underline{a}_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} \uparrow \\ \underline{y} \\ \downarrow \end{bmatrix}$$

it is clear that the equation will have a solution if and only if, $\underline{y}$ is a linear combination of the column vectors of $\underline{A}$; in which case, $\underline{y}$ is said to belong to the range of $\underline{A}$, $\underline{y} \in R(\underline{A})$. In any case, we may decompose $\underline{y}$ into the sum $\underline{y} = \underline{y}_1 + \underline{y}_o$, where $\underline{y}_1 \in R(\underline{A})$ and $\underline{y}_o$ is orthogonal to the column vectors of $\underline{A}$. A vector $\underline{y}_o$ with this property is said to belong to the null space of $\underline{A}'$, $\underline{y}_o \in N(\underline{A}')$, since $\underline{A}'\underline{y}_o = \underline{o}$. $R(\underline{A})$ and $N(\underline{A}')$ are called orthogonal complementary subspaces of m dimensional vector space, E^m -- orthogonal, since any member of one is orthogonal to any member of the other -- complementary, since their union is E^m -- subspaces, since any linear combination of vectors belonging to $R(\underline{A})$ (or $N(\underline{A}')$) also belongs to $R(\underline{A})$ (or $N(\underline{A}')$). The intersection of $R(\underline{A})$ and $N(\underline{A}')$ is the zero vector.

Suppose that S and T are orthogonal complementary subspaces of E^m. Consider a transformation $\underline{P}$ on E^m such that if $\underline{w} \in E^m$ and $\underline{w} = \underline{u} + \underline{v}$, where $\underline{u} \in S$ and $\underline{v} \in T$, then

$$\underline{Pw} = \underline{u}$$

Such a transformation $\underline{P}$ is called the orthogonal projection of E^m onto S.

Necessary and sufficient conditions that $\underline{P}$ be an orthogonal projection are

$$\underline{P}^2 = \underline{P}$$

$$\underline{P}' = \underline{P}$$

We use the notation $\underline{P}_{R(\underline{A})}$ to signify the orthogonal projection onto $R(\underline{A})$.

Let us turn our attention to the diagram of Fig. B-1. This figure should be interpreted in the following manner; any point in the lower plane (E^n) is mapped in the direction of the arrow by the transformation $\underline{A}$ to a point in $R(\underline{A})$ in the upper plane (E^m).

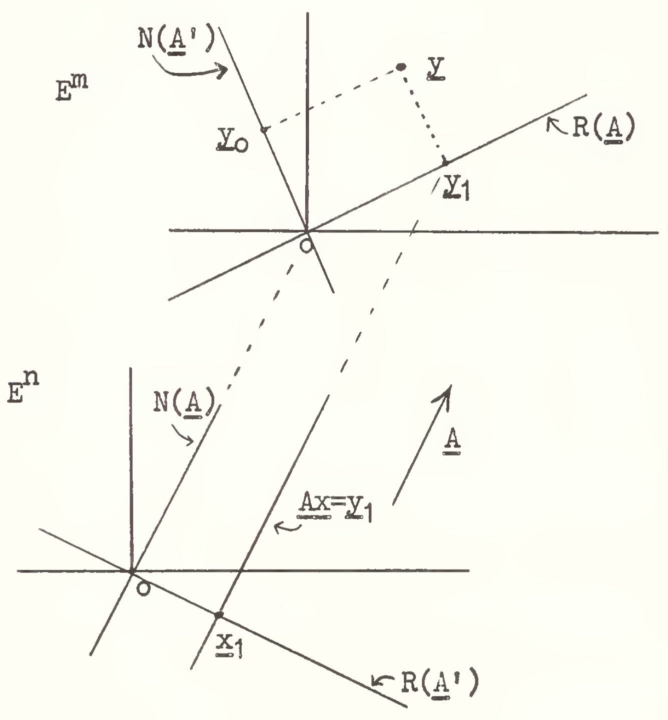


Fig. B-1

Consider the equation $\underline{Ax} = \underline{y}$, where $\underline{y}$ is shown in Fig. B-1. This equation has no solution since $\underline{y} \notin R(\underline{A})$. The point nearest to $\underline{y}$ for which this equation has a solution is $\underline{y}_1$, where $\underline{y}_1 = \underline{P}_{R(\underline{A})}\underline{y}$. If we now consider instead the equation $\underline{Ax} = \underline{y}_1$, we see that any point $\underline{x}$ on the line of Fig. B-1 marked $\underline{Ax} = \underline{y}_1$ is a solution of this equation. From these solutions we choose $\underline{x}_1$, the one closest to the origin. Note that $\underline{x}_1 \in R(\underline{A}')$. We would like to have a transformation $\underline{A}^{\#}$ which would perform these operations. Then $\underline{x}_1$ would be given by

$$\underline{x}_1 = \underline{A}^{\#}\underline{y} \tag{B.30}$$

Let us consider the properties such an $\underline{A}^{\#}$ would possess. First, $R(\underline{A}^{\#}) \subset R(\underline{A}')$ since every point $\underline{y}$ would be mapped by $\underline{A}^{\#}$ into a point belonging to $R(\underline{A}')$. Second, if $\underline{x}_1$ is to be given by (B.30), we would require $\underline{Ax}_1$ to be $\underline{P}_{R(\underline{A})}\underline{y}$. That is, for all $\underline{y}$

$$\underline{AA}^{\#}\underline{y} = \underline{P}_{R(\underline{A})}\underline{y}$$

or

$$\underline{AA}^{\#} = \underline{P}_{R(\underline{A})} \tag{B.31}$$

Third, if $\underline{Ax} = \underline{y}_1$, then $\underline{A}^{\#}\underline{y}_1$ would be $\underline{P}_{R(\underline{A}')}\underline{x}$. That is, for all $\underline{x}$

$$\underline{A}^{\#}\underline{Ax} = \underline{P}_{R(\underline{A}')}\,\underline{x}$$

or

$$\underline{A}^{\#}\underline{A} = \underline{P}_{R(\underline{A}')} \tag{B.32}$$

The condition (B.32) implies that $R(\underline{A}') \subset R(\underline{A}^{\#})$ since, if $\underline{x} \in R(\underline{A}')$, then $\underline{A}^{\#}\underline{A}\underline{x} = \underline{x}$ and $\underline{x} \in R(\underline{A}^{\#})$. We now have the relation $R(\underline{A}^{\#}) \subset R(\underline{A}')$ and $R(\underline{A}') \subset R(\underline{A}^{\#})$ which implies $R(\underline{A}') \equiv R(\underline{A}^{\#})$. Thus, (B.32) becomes

$$\underline{A}^{\#}\underline{A} = \underline{P}_{R(\underline{A}^{\#})} \tag{B.33}$$

The condition (B.31) is equivalent to (B.1) and (B.2), and the condition (B.33) is equivalent to (B.3) and (B.4). Hence, all the properties of the generalized inverse follow from (B.31) and (B.33) and are easily given a geometric interpretation as is done in Fig. B-1.

B6 Further Discussion of the Equation $\underline{A}\underline{x} = \underline{y}$

In this section we shall prove three simple lemmas which will be used repeatedly in Chapter IV. We again consider the linear equation

$$\underline{A}\underline{x} = \underline{y} \tag{B.34}$$

where $\underline{y}$ and $\underline{A}$ are known and $\underline{x}$ is to be determined.

B6.1 Lemma:

The solution of (B.34) is unique if and only if $\underline{A}'\underline{A}$ is positive definite. If $\underline{A}'\underline{A}$ is positive definite the unique solution is

$$\underline{x} = [\underline{A}'\underline{A}]^{-1}\underline{A}'\underline{y} \tag{B.35}$$

Proof: Let $\underline{x}_1$ and $\underline{x}_2$ be any two solutions of (B.34). Then

$$\underline{o} = \underline{A}\,[\underline{x}_1 - \underline{x}_2]$$

and

$$[\underline{x}_1' - \underline{x}_2'] \; \underline{A}'\underline{A} \; [\underline{x}_1 - \underline{x}_2] = 0 \tag{B.36}$$

If $\underline{A}'\underline{A}$ is positive definite (B.36) implies $\underline{x}_1 = \underline{x}_2$ and the solution of (B.34) is unique. If $\underline{A}'\underline{A}$ is not positive definite, let $\underline{x}_o$ be any non-zero vector such that $\underline{x}_o' \underline{A}'\underline{A}\underline{x}_o = 0$, (i.e., $\underline{x}_o \in N(\underline{A})$). If $\underline{x}_1$ is a solution of (B.34) then so is $\underline{x}_1 + \underline{x}_o$ and the solution of (B.34) is not unique. The solution is given by (B.35) since $[\underline{A}'\underline{A}]^{-1}\underline{A}'$ is the generalized inverse of A.

B6.2 Lemma:

A solution of (B.34) will exist if and only if $\underline{y} \in R(\underline{AA}')$; in which case the solution of smallest Euclidian norm is given by

$$\underline{x} = \underline{A}'(\underline{AA}')^{\#}\underline{y} \tag{B.37}$$

Proof: Using the theorem of section B3 and recalling that $\underline{A}'(\underline{AA}')^{\#} = \underline{A}^{\#}$, we observe that (B.34) will have a solution if and only if

$$\underline{y} = \underline{AA}^{\#}\underline{y} = \underline{AA}'(\underline{AA}')^{\#}\underline{y}$$

Hence, (B.34) will have a solution only if $\underline{y} \in R(\underline{AA}')$. Moreover, by the same theorem, $\underline{y} = \underline{AA}'\underline{\lambda}$ has a solution, (i.e., $\underline{y} \in R(\underline{AA}')$, if and only if

$$\underline{y} = \underline{AA}'(\underline{AA}')^{\#}\underline{y}$$

The solution of smallest Euclidian norm is given by (B.37) since $\underline{A}'(\underline{AA}')^{\#} = \underline{A}^{\#}$.

B6.3 Lemma:

A solution of (B.34) will exist for all $\underline{y}$ if and only if $\underline{AA}'$ is positive definite; in which case the solution of smallest Euclidian norm is given by

$$\underline{x} = \underline{A}'(\underline{AA}')^{-1}\underline{y}$$

Proof: A solution of (B.34) will exist for all $\underline{y}$ if and only if all $\underline{y} \in R(\underline{A})$, or equivalently if and only if $N(\underline{A}') = \underline{o}$. But a vector $\underline{y}_o$ belongs to $N(\underline{A}')$ if and only if $\underline{y}_o' \underline{AA}'\underline{y}_o = o$. Thus, $N(\underline{A}') = \underline{o}$ if and only if $\underline{AA}'$ is positive definite.

APPENDIX C

GAUSSIAN RANDOM VECTORS

In this appendix we summarize some known facts about Gaussian random vectors or, equivalently, about the multivariate normal distribution. Special attention is given to situations in which the covariance matrix is singular. The pseudo-inverse and generalized inverse of Appendix B are shown to be applicable to these situations.

If $\underline{x}$ is an n-dimensional random vector with mean $\underline{m}$ and covariance matrix $\underline{V} = E\left[(\underline{x}-\underline{m})(\underline{x}-\underline{m})'\right]$, we shall say that $\underline{x}$ is a Gaussian random vector if its characteristic function is of the following form;

$$C_{\underline{x}}(i\underline{w}) = E\left[\exp i\underline{w}'\underline{x}\right] = \exp\left[i\underline{w}'\underline{m} - \tfrac{1}{2}\ \underline{w}'\underline{V}\underline{w}\right] \tag{C.1}$$

If $\underline{V}$ is non-singular (positive definite) the probability density function for $\underline{x}$ takes the following form;

$$f(\underline{x}) = \frac{1}{(2\pi)^{n/2}\ |\underline{V}|^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}\|\underline{x}-\underline{m}\|^{2}_{\underline{V}^{-1}}\right] \tag{C.2}$$

The definition of a Gaussian random vector in terms of the characteristic function has been suggested [29] in order to avoid the difficulties which arise when $\underline{V}$ is singular. The results to be presented in this appendix have been derived elsewhere [29] using a different approach based on the characteristic function. We shall use an approach which is more in keeping with the body of this work.

A vector $\underline{u}$ will be called a normalized Gaussian random vector if it is Gaussian with mean $E(\underline{u}) = \underline{o}$ and covariance matrix $E(\underline{u}\underline{u}') = \underline{I}$. Any

Gaussian random vector $\underline{x}$ can be considered as having been obtained from a normalized Gaussian random vector of the same dimension by a linear transformation of the following form;

$$\underline{x}-\underline{m} = \underline{Tu}$$

Then,

$$E(\underline{x}) = \underline{m} + \underline{T}\, E(\underline{u}) = \underline{m}$$

$$E\left[(\underline{x}-\underline{m})(\underline{x}-\underline{m}')\right] = \underline{T}\, E(\underline{uu}')\, \underline{T}' = \underline{TT}' = \underline{V}$$

It is shown in Appendix A that a non-negative definite matrix $\underline{V}$ may always be factored into a product $\underline{TT}'$. It will suffice for our purpose to know that the transformation $\underline{T}$ exists, since we shall never have to compute $\underline{T}$.

Let $\underline{x}_1$ and $\underline{x}_2$ be Gaussian random vectors having the joint distribution specified by

$$\underline{x} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \end{bmatrix} \qquad \underline{m} = \begin{bmatrix} \underline{m}_1 \\ \underline{m}_2 \end{bmatrix} \qquad \underline{V} = \begin{bmatrix} \underline{V}_{11} & \underline{V}_{12} \\ \underline{V}_{21} & \underline{V}_{22} \end{bmatrix}$$

where $\underline{V}$ may be singular.

We consider $\underline{x}$ to be the result of a transformation on a normalized Gaussian vector $\underline{u}$. Then,

$$\underline{x}-\underline{m} = \underline{Tu} \quad , \quad \underline{TT}' = \underline{V} \tag{C.3}$$

Let $\underline{H}$ be the matrix $[\underline{0}\ \underline{I}]$ so that $\underline{x}_2 = \underline{Hx}$. Suppose that an experiment is performed in which it is learned that the value of $\underline{x}_2$ is $\underline{z}$. Having learned the value of $\underline{x}_2$ we want to know the *a posteriori* distribution

for $\underline{x}_1$. The a posteriori distribution for $\underline{x}_1$ will be Gaussian and hence, will be specified by its mean and its covariance matrix. To determine the a posteriori mean we minimize

$$\tfrac{1}{2}\|\underline{u}\|^2 + \underline{\lambda}'(\underline{z}-\underline{Hm}-\underline{HTu}) \tag{C.4}$$

That is, we locate the mode (or mean) of the a posteriori distribution for $\underline{u}$. The Lagrange multiplier has been introduced to insure that $\underline{x}_2$ takes on the observed value $\underline{z}$. Minimizing (C.4) is exactly the same type of problem as was discussed in Appendix B. Proceeding as before, we set the partial derivatives, with respect to $\underline{u}$ and $\underline{\lambda}$, of (C.4) equal to zero to obtain the following two equations for the minimizing value $\underline{u}$;

$$\underline{HT}\hat{\underline{u}} = \underline{z}-\underline{Hm} \tag{C.5}$$

$$\hat{\underline{u}} = \underline{T}'\underline{H}'\underline{\lambda} \tag{C.6}$$

Combining (C.5) and (C.6) we obtain

$$\underline{HTT}'\underline{H}'\,\underline{\lambda} = \underline{z}-\underline{Hm} \tag{C.7}$$

As was shown in Appendix B, a $\underline{\lambda}$ satisfying (C.7) will always exist if there is a $\hat{\underline{u}}$ which satisfies (C.5). There will be a $\hat{\underline{u}}$ which satisfies (C.5) unless the observed value of $\underline{x}_2$ has probability zero; that is, unless the singularity of the a priori covariance matrix $\underline{V}$ makes the observed value inconsistent with prior knowledge. We rule out this possibility since it leads to division by zero in Bayes' Theorem and violates our basic assumption. Values of $\underline{x}_2$ for which a solution to (C.5) exists will be called consistent. Substituting for $\underline{H}$ and $\underline{TT}'$ in (C.7) yields

$$\underline{V}_{22}\,\underline{\lambda} = \underline{z}-\underline{m}_2 \tag{C.8}$$

Any pseudo-inverse may be used to solve for $\underline{\lambda}$. Then,

$$\underline{\lambda} = \underline{V}_{22}^{\dagger}\,(\underline{z}-\underline{m}_2) \tag{C.9}$$

Combining (C.9) and (C.6) gives the following equation;

$$\hat{\underline{u}} = \underline{T}'\underline{H}'\underline{V}_{22}^{\dagger}(\underline{z}-\underline{m}_2)$$

Finally, using (C.3), we obtain

$$\hat{\underline{x}}-\underline{m} = \underline{VH}'\underline{V}_{22}^{\dagger}(\underline{z}-\underline{m}_2) \tag{C.10}$$

This is equivalent to the two equations

$$\hat{\underline{x}}_1 = \underline{m}_1 + \underline{V}_{12}\underline{V}_{22}^{\dagger}(\underline{z}-\underline{m}_2) \tag{C.11}$$

$$\hat{\underline{x}}_2 = \underline{m}_2 + \underline{V}_{22}\underline{V}_{22}^{\dagger}(\underline{z}-\underline{m}_2) \tag{C.12}$$

From the pseudo-inverse theorem of Appendix B we know that the existence of a solution of (C.8) implies the following relation;

$$\underline{z}-\underline{m}_2 = \underline{V}_{22}\underline{V}_{22}^{\dagger}(\underline{z}-\underline{m}_2) \tag{C.13}$$

Combining (C.12) and (C.13) leads us to the consistent result $\underline{x}_2 = \underline{z}$. The basic result is (C.11) and is summarized in the following statement.

<u>Statement C-1</u>: The conditional expectation of $\underline{x}_1$, given a consistent $\underline{x}_2$ is given by the following relation;

$$\hat{\underline{x}}_1 = \underline{m}_1 + \underline{V}_{12}\underline{V}_{22}^{\dagger}(\underline{x}_2-\underline{m}_2)$$

We may obtain a useful relation from (C.13) by using the fact that for all consistent values of $\underline{x}_2$ the following relation is satisfied;

$$\underline{x}_2-\underline{m}_2 = \underline{V}_{22}\underline{V}_{22}^{\dagger}(\underline{x}_2-\underline{m}_2) \tag{C.14}$$

Postmultiplying both sides of (C.14) by $(\underline{x}_1-\underline{m}_1)'$ and averaging with respect to the joint a priori distribution yields

$$\underline{V}_{21} = \underline{V}_{22}\underline{V}_{22}^{\dagger}\underline{V}_{21} \tag{C.15}$$

Taking the transpose of (C.15) gives the following useful relation;

$$\underline{V}_{12} = \underline{V}_{12}\underline{V}_{22}^{\dagger}\underline{V}_{22} \tag{C.16}$$

In the development of an expression for the covariance matrix of the a posteriori distribution for $\underline{x}_1$, given $\underline{x}_2$, the concept of pre-posterior analysis [54] will be used. Given any value of $\underline{x}_2$ we may find the a posteriori distribution for $\underline{x}_1$ which will, in general, depend on the value of $\underline{x}_2$. Before the value of $\underline{x}_2$ is known, parameters of the a posteriori distribution may be considered to be random variables. Pre-posterior analysis is concerned with the a priori probability of obtaining particular values for parameters of the a posteriori distribution. We shall need the following simple lemma;

Lemma: Let $\underset{\underline{x}_1|\underline{x}_2}{E}[\phi(\underline{x}_1)]$ be the expected value of some function $\phi(\underline{x}_1)$ with respect to the a posteriori distribution. Then the expected value of $\underset{\underline{x}_1|\underline{x}_2}{E}[\phi(\underline{x}_1)]$ with respect to the a priori distribution for $\underline{x}_2$ is the a priori expected value of $\phi(\underline{x}_1)$. That is,

$$\underset{\underline{x}_2}{E}\left[\underset{\underline{x}_1|\underline{x}_2}{E}[\phi(\underline{x}_1)]\right] = \underset{\underline{x}_1}{E}[\phi(\underline{x}_1)]$$

The proof is by inspection since on the left side of this expression

the average is with respect to the joint a priori distribution for $\underline{x}_1$ and $\underline{x}_2$.

Using this lemma we may easily prove the following theorem given in [54].

Theorem: The expected value with respect to $\underline{x}_2$ of the covariance matrix of the a posteriori distribution plus the covariance matrix with respect to $\underline{x}_2$ of the mean of the a posteriori distribution is equal to the covariance matrix of the a priori distribution for $\underline{x}_1$. That is,

$$\underset{\underline{x}_2}{E}\left\{\underset{\underline{x}_1|\underline{x}_2}{E}[\underline{x}_1\underline{x}_1'] - \underset{\underline{x}_1|\underline{x}_2}{E}(\underline{x}_1)\left[\underset{\underline{x}_1|\underline{x}_2}{E}(\underline{x}_1)\right]'\right\}$$

$$+ \underset{\underline{x}_2}{E}\left\{\underset{\underline{x}_1|\underline{x}_2}{E}(\underline{x}_1)\left[\underset{\underline{x}_1|\underline{x}_2}{E}(\underline{x}_1)\right]'\right\} - \underset{\underline{x}_2}{E}\left[\underset{\underline{x}_1|\underline{x}_2}{E}(\underline{x}_1)\right]\left\{\underset{\underline{x}_2}{E}\left[\underset{\underline{x}_1|\underline{x}_2}{E}(\underline{x}_1)\right]\right\}'$$

$$= \underset{\underline{x}_1}{E}[\underline{x}_1\underline{x}_1'] - \underset{\underline{x}_1}{E}[\underline{x}_1]\left[\underset{\underline{x}_1}{E}[\underline{x}_1]\right]'$$

The proof is immediate by canceling like terms and applying the lemma. This theorem is sometimes stated for scalar random variables in the following concise form. The mean of the posterior variance plus the variance of the posterior mean is equal to the prior variance.

Statement C-2: The covariance matrix with respect to $\underline{x}_2$ of the mean of the a posteriori distribution for $\underline{x}_1$ is $\underline{V}_{12}\underline{V}_{22}^{\dagger}\underline{V}_{21}$. This statement follows from C-1 since

$$\underset{\underline{x}_2}{E}[(\hat{\underline{x}}_1-\underline{m})(\hat{\underline{x}}_1-\underline{m}_1)'] = \underline{V}_{12}\underline{V}_{22}^{\dagger}\underline{V}_{22}\underline{V}_{22}^{\dagger}\underline{V}_{21}$$

Using (C.15) or (C.16) concludes the proof.

Statement C-3: $\underline{x}_1 - \underset{\underline{x}_1|\underline{x}_2}{E}(\underline{x}_1)$ is independent of $\underline{x}_2$.

To show this we compute the covariance matrix of this quantity and $\underline{x}_2$. From Statement C-1 we have

$$\underset{\underline{x}_1|\underline{x}_2}{E}\left[(\underline{x}_1 - \underset{\underline{x}_1|\underline{x}_2}{E}(\underline{x}_1))(\underline{x}_2-\underline{m}_2)'\right]$$

$$= \underset{\underline{x}_1|\underline{x}_2}{E}\left\{[\underline{x}_1-\underline{m}_1 - \underline{V}_{12}\underline{V}_{22}^{\dagger}(\underline{x}_2-\underline{m}_2)]\quad(\underline{x}_2-\underline{m}_2)'\right\} = \underline{V}_{12} - \underline{V}_{12}\underline{V}_{22}^{\dagger}\underline{V}_{22}$$

This quantity is equal to $\underline{0}$ by (C.16).

This statement implies that the covariance matrix of the a posteriori distribution for $\underline{x}_1$ is independent of the observed value of $\underline{x}_2$ and hence, the expected value with respect to $\underline{x}_2$ of this quantity is equal to the quantity itself. Using this fact and the theorem leads to the following statement.

Statement C-4: The covariance matrix of the a posteriori distribution for $\underline{x}_1$ given $\underline{x}_2$ is $\underline{V}_{11} - \underline{V}_{12}\underline{V}_{22}^{\dagger}\underline{V}_{21}$.

The generalized inverse may, of course, always be used in place of the pseudo-inverse. The use of the pseudo-inverse has been consistent with our effort to use a notation compatible with [29]. In practice one might prefer to use the unique generalized inverse.

APPENDIX D

GENERAL SOLUTION OF THE DISCRETE-TIME LINEAR ESTIMATION PROBLEM

D-1 Introduction

This appendix is devoted to a study of a very general form of the linear discrete-time estimation problem in which the covariance matrices may be singular. For the sake of generality and at the sacrifice of some simplicity, we allow correlation between the random input $\underline{w}$ and the measurement noise $\underline{v}$. Necessary background material appears in Chapter II and Appendices B and C.

D-2 Problem Statement

Consider the system

$$\underline{x}(k+1) = \underline{F}(k)\underline{x}(k) + \underline{G}(k)\underline{w}(k) \tag{D.1}$$

$$\underline{z}(k) = \underline{H}(k)\underline{x}(k) + \underline{v}(k) \tag{D.2}$$

where $\underline{x}$ is the state vector and $\underline{z}$ is the observation. $\{\underline{w}(k)\}$, the random input and $\{\underline{v}(k)\}$, the measurement noise, are sequences of independent random variables with zero means and the following covariance matrices;

$$\left.\begin{aligned} E\left[\underline{w}(j)\ \underline{w}'(k)\right] &= \underline{Q}(k)\ \delta_{jk} \\ E\left[\underline{v}(j)\ \underline{v}'(k)\right] &= \underline{R}(k)\ \delta_{jk} \\ E\left[\underline{w}(j)\ \underline{v}'(k)\right] &= \underline{S}(k)\ \delta_{jk} \end{aligned}\right\} \quad \text{for all intergers j and k}$$

Based on a finite observation sequence $\{\underline{z}(o),\ldots,\underline{z}(n)\}$, we seek an estimate of the sequence of states $\{\underline{x}(o),\ldots,\underline{x}(n+m)\}$.

D3 Filtering and Prediction

Suppose that, given the observed sequence $\{\underline{z}(o),\ldots,\underline{z}(k-1)\}$, the distribution for $\underline{x}(k)$ has mean $\hat{\underline{x}}(k|k-1)$ and covariance matrix $\underline{P}(k)$. Consider the vector

$$\begin{bmatrix} \underline{x}(k) \\ \underline{z}(k) \end{bmatrix} = \begin{bmatrix} \underline{x}(k) \\ \underline{H}(k)\underline{x}(k) + \underline{v}(k) \end{bmatrix} \quad \text{with mean} \quad \begin{bmatrix} \hat{\underline{x}}(k|k-1) \\ \underline{H}(k)\hat{\underline{x}}(k|k-1) \end{bmatrix}$$

and having the following covariance matrix;

$$\begin{bmatrix} \underline{P}(k) & \underline{P}(k)\underline{H}'(k) \\ \underline{H}(k)\underline{P}(k) & \underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k) \end{bmatrix}$$

If $\underline{z}(k)$ is observed, we may use statements (C-1) and (C-4) from Appendix C to obtain the following expressions for the mean $\hat{\underline{x}}(k|k)$ and covariance matrix $\underline{C}(k)$ of the distribution for $\underline{x}(k)$ given $\{\underline{z}(o),\ldots,\underline{z}(k)\}$;

$$\hat{\underline{x}}(k|k) = \hat{\underline{x}}(k|k-1) + \underline{P}(k)\underline{H}'(k)\left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)\right]^{\dagger} \left[\underline{z}(k)-\underline{H}(k)\hat{\underline{x}}(k|k-1)\right] \qquad (D.3)$$

$$\underline{C}(k) = \underline{P}(k)-\underline{P}(k)\underline{H}'(k)\left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)\right]^{\dagger} \underline{H}(k)\underline{P}(k) \qquad (D.4)$$

Similarly, consider the vector

$$\begin{bmatrix} \underline{x}(k+1) \\ \underline{z}(k) \end{bmatrix} = \begin{bmatrix} \underline{F}(k)\underline{x}(k) + \underline{G}(k)\underline{w}(k) \\ \underline{H}(k)\underline{x}(k) + \underline{v}(k) \end{bmatrix} \text{with mean} \quad \begin{bmatrix} \underline{F}(k)\hat{\underline{x}}(k|k-1) \\ \underline{H}(k)\hat{\underline{x}}(k|k-1) \end{bmatrix}$$

and having the following covariance matrix;

$$\begin{bmatrix} \underline{F}(k)\underline{P}(k)\underline{F}'(k) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k) & \underline{F}(k)\underline{P}(k)\underline{H}'(k) + \underline{G}(k)\underline{S}(k) \\ \underline{H}(k)\underline{P}(k)\underline{F}'(k) + \underline{S}'(k)\underline{G}'(k) & \underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k) \end{bmatrix}$$

If $\underline{z}(k)$ is observed we may again use statements (C-1) and (C-4) to obtain the following expressions for the mean $\hat{\underline{x}}(k+1|k)$ and covariance matrix $\underline{P}(k+1)$ of the distribution for $\underline{x}(k+1)$ given $\{\underline{z}(o),\ldots,\underline{z}(k)\}$;

$$\hat{\underline{x}}(k+1|k) = \underline{F}(k)\hat{\underline{x}}(k|k-1) + [\underline{F}(k)\underline{P}(k)\underline{H}'(k) + \underline{G}(k)\underline{S}(k)]$$
$$[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)]^{\dagger}[\underline{z}(k)-\underline{H}(k)\hat{\underline{x}}(k|k-1)] \quad (D.5)$$

$$\underline{P}(k+1) = \underline{F}(k)\underline{P}(k)\underline{F}'(k) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k)$$
$$-[\underline{F}(k)\underline{P}(k)\underline{H}'(k) + \underline{G}(k)\underline{S}(k)]\ [\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)]^{\dagger}$$
$$[\underline{H}(k)\underline{P}(k)\underline{F}'(k) + \underline{S}'(k)\underline{G}'(k)] \quad (D.6)$$

With practically no effort we have obtained the solution to the linear filtering problem and the solution to the problem of predicting one step ahead. Equations (D.5) and (D.6) are, except for slight details, the same ones as those given in [29].

Note that $\hat{\underline{x}}(k+1|k)$ is not simply an extrapolation of $\hat{\underline{x}}(k|k)$. This is due to the fact that information about $\underline{w}(k)$ is available in $\underline{z}(k)$ because $\underline{w}(k)$ and $\underline{v}(k)$ are correlated. However, $\hat{\underline{x}}(k+2|k)$ and predictions further into the future are simply extrapolations of $\hat{\underline{x}}(k+1|k)$, since no information is available about future values of $\underline{w}$. Since it is now evident how to make predictions, we may confine ourselves to the problem of estimating $\{\underline{x}(o),\ldots,\underline{x}(n+1)\}$ given $\{\underline{z}(o),\ldots,\underline{z}(n)\}$.

D4 Preliminary Considerations

Let $\underline{r}(k) = \underline{G}(k)\underline{w}(k)$, and consider the vector

$$\begin{bmatrix} \underline{r}(k) \\ \underline{v}(k) \end{bmatrix} \quad \text{with covariance matrix} \quad \begin{bmatrix} \underline{G}(k)\underline{Q}(k)\underline{G}'(k) & \underline{G}(k)\underline{S}(k) \\ \underline{S}'(k)\underline{G}'(k) & \underline{R}(k) \end{bmatrix}$$

We consider the vector formed by combining $\underline{r}$ and $\underline{v}$ to have been obtained by a transformation on a normalized vector $\underline{u}(k)$, $E[\underline{u}(j)\underline{u}'(k)] = \underline{I}\,\delta_{jk}$.

$$\begin{bmatrix} \underline{r}(k) \\ \underline{v}(k) \end{bmatrix} = \underline{T}(k)\underline{u}(k), \qquad \underline{T}(k)\underline{T}'(k) = \begin{bmatrix} \underline{G}(k)\underline{Q}(k)\underline{G}'(k) & \underline{G}(k)\underline{S}(k) \\ \underline{S}'(k)\underline{G}'(k) & \underline{R}(k) \end{bmatrix}$$

Let us introduce the matrices $\underline{V} = [\underline{0},\underline{I}]$ and $\underline{W} = [\underline{I},\underline{0}]$ such that

$$\underline{r}(k) = \underline{WT}(k)\underline{u}(k) \tag{D.7}$$

$$\underline{v}(k) = \underline{VT}(k)\underline{u}(k) \tag{D.8}$$

$$\underline{WT}(k)\underline{T}'(k)\underline{W}' = \underline{G}(k)\underline{Q}(k)\underline{G}'(k) \tag{D.9}$$

$$\underline{VT}(k)\underline{T}'(k)\underline{V}' = \underline{R}(k) \tag{D.10}$$

$$\underline{WT}(k)\underline{T}'(k)\underline{V}' = \underline{G}(k)\underline{S}(k) \tag{D.11}$$

The a priori distribution for the initial state is Gaussian with mean $\underline{m}$ and covariance matrix $\underline{P}(o)$ which may be singular. We shall consider the initial state as having been obtained by a transformation on a normalized vector $\underline{\theta}$

$$\underline{x}(o) - \underline{m} = \underline{A\theta} \qquad\qquad \underline{AA}' = \underline{P}(o) \tag{D.12}$$

D5 Derivation of the General Solution

Introducing the Lagrange multiplier vectors $\underline{\delta}$, $\underline{\lambda}(k)$ and $\underline{\varrho}(k)$, we may minimize the following expression;

$$I_{n,1} = \tfrac{1}{2}\|\underline{\theta}\|^2 + \underline{\delta}'[\underline{x}(o)-\underline{m}-\underline{A\theta}] + \sum_{k=o}^{n}\Big\{\underline{\varrho}'(k)\,[\underline{z}(k)-\underline{H}(k)\underline{x}(k)-\underline{VT}(k)\underline{u}(k)]$$

$$+ \tfrac{1}{2}\|\underline{u}(k)\|^2 + \underline{\lambda}'(k)\,[\underline{x}(k+1)-\underline{F}(k)\underline{x}(k)-\underline{WT}(k)\underline{u}(k)]\Big\} \tag{D.13}$$

Setting equal to zero the partial derivatives of $I_{n,1}$, with respect to $\underline{u}(k)$, $\underline{\varrho}(k)$, $\underline{\lambda}(k)$ (for k=o,...,n); $\underline{x}(k)$ (for k=o,...,n+1); and θ, yields the following relations;

$$\underline{\theta} = \underline{A}'\underline{\beta} \tag{D.14}$$

$$\underline{\beta} = \underline{H}'(o)\underline{\varrho}(o) + \underline{F}'(o)\underline{\lambda}(o) \tag{D.15}$$

$$\underline{u} = \underline{T}'(k)\underline{V}'\underline{\varrho}(k) + \underline{T}'(k)\underline{W}'\underline{\lambda}(k) \tag{D.16}$$

$$\underline{\lambda}(k-1) = \underline{F}'(k)\underline{\lambda}(k) + \underline{H}'(k)\underline{\varrho}(k) \qquad k=1,\ldots,n \tag{D.17}$$

$$\underline{\lambda}(n) = \underline{o} \tag{D.18}$$

$$\hat{\underline{x}}(o|n) = \underline{m} + \underline{A\theta} \tag{D.19}$$

$$\hat{\underline{x}}(k+1|n) = \underline{F}(k)\hat{\underline{x}}(k|n) + \underline{WT}(k)\underline{u}(k) \qquad k=o,\ldots,n \tag{D.20}$$

$$\underline{z}(k) = \underline{H}(k)\hat{\underline{x}}(k|n) + \underline{VT}(k)\underline{u}(k) \tag{D.21}$$

Combining (D.14), (D.15) and (D.19) and observing the $\underline{AA}' = \underline{P}(o)$ yields

$$\hat{\underline{x}}(o|n) = \underline{m} + \underline{P}(o)\underline{H}'(o)\underline{\varrho}(o) + \underline{P}(o)\underline{F}'(o)\underline{\lambda}(o) \tag{D.22}$$

Combining (D.16) and (D.20) and using (D.9) and (D.11), we obtain the following equation;

$$\hat{\underline{x}}(k+1|n) = \underline{F}(k)\hat{\underline{x}}(k|n) + \underline{G}(k)\underline{S}(k)\underline{\varrho}(k) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k)\underline{\lambda}(k) \tag{D.23}$$

Using (D.10), (D.11) and (D.16), we rewrite (D.21) in the following form;

$$\underline{z}(k) = \underline{H}(k)\hat{\underline{x}}(k|n) + \underline{R}(k)\underline{\varrho}(k) + \underline{S}'(k)\underline{G}'(k)\underline{\lambda}(k) \tag{D.24}$$

We now have a two-point boundary value problem given by (D.23) and (D.17) and the constraint (D.24). The boundary conditions are given by (D.18) and (D.22).

In order to simplify future equations we define the following quantities;

$$\underline{B}(k) = \left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)\right]^{\#} \tag{D.25}$$

$$\underline{M}(k) = \underline{B}(k) \left[\underline{H}(k)\underline{P}(k)\underline{F}'(k) + \underline{S}'(k)\underline{G}'(k)\right] \tag{D.26}$$

We begin by considering the boundary condition for $\hat{\underline{x}}(o|n)$. Combining (D.22) and (D.24) yields

$$\begin{aligned}\underline{z}(o) = \underline{H}(o)\underline{m} + &\left[\underline{H}(o)\underline{P}(o)\underline{H}'(o) + \underline{R}(o)\right] \underline{\varrho}(o) \\ &+ \left[\underline{H}(o)\underline{P}(o)\underline{F}'(o) + \underline{S}'(o)\underline{G}'(o)\right] \underline{\lambda}(o)\end{aligned} \tag{D.27}$$

Using the generalized inverse to solve for a $\underline{\varrho}(o)$ yields

$$\underline{\varrho}(o) = \underline{B}(o) \left[\underline{z}(o)-\underline{H}(o)\underline{m}\right] - \underline{M}(o)\underline{\lambda}(o) \tag{D.28}$$

Note that any pseudo-inverse could have been used to solve for $\underline{\varrho}(o)$. The generalized inverse is convenient because it allows us to define $\underline{B}(k)$ uniquely. Combining (D.28) and (D.22) yields

$$\begin{aligned}\hat{\underline{x}}(o|n) = \underline{m} + &\underline{P}(o)\underline{H}'(o)\underline{B}(o) \left[\underline{z}(o)-\underline{H}(o)\underline{m}\right] \\ &+ \left[\underline{P}(o)\underline{F}'(o)-\underline{P}(o)\underline{H}'(o)\underline{M}(o)\right] \underline{\lambda}(o)\end{aligned} \tag{D.29}$$

or, equivalently,

$$\hat{\underline{x}}(o|n) = \hat{\underline{x}}(o|o) + \left[\underline{P}(o)\underline{F}'(o)-\underline{P}(o)\underline{H}'(o)\underline{M}(o)\right] \underline{\lambda}(o) \tag{D.30}$$

It is reassuring to note that for the special case $\underline{S} = \underline{0}$, (D.30) reduces to the following familiar equation of Chapter II;

$$\hat{\underline{x}}(o|n) = \hat{\underline{x}}(o|o) + \underline{C}(o)\underline{F}'(o)\underline{\lambda}(o) \qquad ,\underline{S}(o) = \underline{0} \tag{2.41}$$

We now hypothesize that the general solution is of the following form;

$$\underline{\hat{x}}(k|n) = \underline{\hat{x}}(k|k) + [\underline{P}(k)\underline{F}'(k) - \underline{P}(k)\underline{H}'(k)\underline{M}(k)]\,\underline{\lambda}(k) \quad ,k=0,\ldots,n \qquad (D.31)$$

$$\underline{\lambda}(k-1) = [\underline{F}'(k)-\underline{H}'(k)\underline{M}(k)]\;\underline{\lambda}(k) + \underline{H}'(k)\underline{B}(k)\,[\underline{z}(k)-\underline{H}(k)\underline{\hat{x}}(k|k-1)] \qquad (D.32)$$

where $\underline{\hat{x}}(k|k)$, $\underline{\hat{x}}(k+1|k)$ and $\underline{P}(k)$ are given by (D.3), (D.5) and (D.6). The hypothesis (D.32) is equivalent to requiring that $\underline{\rho}(k)$ be given by the following equation;

$$\underline{\rho}(k) = \underline{B}(k)\;[\underline{z}(k)-\underline{H}(k)\underline{\hat{x}}(k|k-1)] - \underline{M}(k)\underline{\lambda}(k) \qquad (D.33)$$

We have already shown that (D.31) and (D.33) are satisfied for k=o. To establish a proof by induction we need to show that if (D.31) and (D.33) are satisfied for some $k \in (o,\ldots,n-1)$, then (D.31) and (D.33) are satisfied for k+1.

Suppose that (D.31) and (D.33) are satisfied for some $k \in (o,\ldots,n-1)$; then from (D.23) we have the following relation;

$$\begin{aligned}\underline{\hat{x}}(k+1|n) = {} & \underline{F}(k)\underline{\hat{x}}(k|k) + [\underline{F}(k)\underline{P}(k)\underline{F}'(k) - \underline{F}(k)\underline{P}(k)\underline{H}'(k)\underline{M}(k) \\ & + \underline{G}(k)\underline{Q}(k)\underline{G}'(k) - \underline{G}(k)\underline{S}(k)\underline{M}(k)]\,\underline{\lambda}(k) \\ & + \underline{G}(k)\underline{S}(k)\underline{B}(k)\;[\underline{z}(k) - \underline{H}(k)\underline{\hat{x}}(k|k-1)]\end{aligned} \qquad (D.34)$$

Using (D.3), (D.5) and (D.6), we may rewrite (D.34) in the following form;

$$\underline{\hat{x}}(k+1|n) = \underline{\hat{x}}(k+1|k) + \underline{P}(k+1)\,\underline{\lambda}(k) \qquad (D.35)$$

Substituting for $\underline{\lambda}(k)$ from (D.17) yields

$$\underline{\hat{x}}(k+1|n) = \underline{\hat{x}}(k+1|k) + \underline{P}(k+1)\underline{H}'(k+1)\underline{\rho}(k+1) + \underline{P}(k+1)\underline{F}'(k+1)\underline{\lambda}(k+1) \quad (D.36)$$

From (D.24) we may write

$$\underline{z}(k+1) = \underline{H}(k+1)\underline{\hat{x}}(k+1|k) + [\underline{H}(k+1)\underline{P}(k+1)\underline{H}'(k+1) + \underline{R}(k+1)]\ \underline{\rho}(k+1)$$

$$+ [\underline{H}(k+1)\underline{P}(k+1)\underline{F}'(k+1) + \underline{S}'(k+1)\underline{G}'(k+1)]\ \underline{\lambda}(k+1) \qquad (D.37)$$

Using the generalized inverse to solve for a $\underline{\rho}(k+1)$ we obtain the following relation which satisfies the hypothesis, (D.37);

$$\underline{\rho}(k+1) = \underline{B}(k+1)[\underline{z}(k+1)-\underline{H}(k+1)\underline{\hat{x}}(k+1|k)] - \underline{M}(k+1)\underline{\lambda}(k+1)$$

Combining (D.38) and (D.36) and using (D.3), we obtain the following equation which satisfies the hypothesis (D.31) and completes the proof;

$$\underline{\hat{x}}(k+1|n) = \underline{\hat{x}}(k+1|k+1) + [\underline{P}(k+1)\underline{F}'(k+1)-\underline{P}(k+1)\underline{H}'(k+1)\underline{M}(k+1)]\ \underline{\lambda}(k+1) \qquad (D.39)$$

We may summarize these results in the following theorem;

<u>Theorem</u>: The general solution of the linear estimation problem as specified by the two-point boundary value problem (D.23), (D.17), (D.24), (D.18) and (D.22) is given by the following recurrence relations;

$$\underline{\hat{x}}(k+1|k) = \underline{F}(k)\underline{\hat{x}}(k|k-1) + \underline{M}'(k)[\underline{z}(k)-\underline{H}(k)\underline{\hat{x}}(k|k-1)] \qquad (D.40)$$

$$\underline{P}(k+1) = \underline{F}(k)\underline{P}(k)\underline{F}'(k) + \underline{G}(k)\underline{Q}(k)\underline{G}'(k)-\underline{M}'(k)\underline{B}^{\#}(k)\underline{M}(k) \qquad (D.41)$$

$$\underline{\hat{x}}(k|k) = \underline{\hat{x}}(k|k-1) + \underline{P}(k)\underline{H}'(k)\underline{B}(k)[\underline{z}(k)-\underline{H}(k)\underline{\hat{x}}(k|k-1)] \qquad (D.42)$$

$$\underline{\hat{x}}(k|n) = \underline{\hat{x}}(k|k) + [\underline{P}(k)\underline{F}'(k)-\underline{P}(k)\underline{H}'(k)\underline{M}(k)]\ \underline{\lambda}(k) \qquad k=o,\ldots,n \qquad (D.43)$$

$$\underline{\lambda}(k-1) = [\underline{F}'(k)-\underline{H}'(k)\underline{M}(k)]\ \underline{\lambda}(k) + \underline{H}'(k)\underline{B}(k)[\underline{z}(k)-\underline{H}(k)\underline{\hat{x}}(k|k-1)] \qquad (D.44)$$

$$k=1,\ldots,n$$

where $\underline{B}(k)$ and $\underline{M}(k)$ are defined in (D.25) and (D.26). The recursion begins by replacing $\hat{\underline{x}}(k|k-1)$ for k=o by $\underline{m}$ in (D.41) and (D.42). Predictions are extrapolations of $\hat{\underline{x}}(n+1|n)$; for example,

$$\hat{\underline{x}}(n+2|n) = \underline{F}(n+1)\hat{\underline{x}}(n+1|n)$$

$$\hat{\underline{x}}(n+3|n) = \underline{F}(n+2)\hat{\underline{x}}(n+2|n) \tag{D.45}$$

For the special case $\underline{S}(k) = \underline{0}$ for all k (D.43) becomes

$$\hat{\underline{x}}(k|n) = \hat{\underline{x}}(k|k) + \underline{C}(k)\underline{F}'(k)\underline{\lambda}(k) \tag{D.46}$$

where $\underline{C}(k)$ is given by (D.4), and (D.44) beomes

$$\underline{\lambda}(k-1) = \left\{ \underline{I}-\underline{H}'(k)\left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)\right]^{\#}\underline{H}(k)\underline{P}(k) \right\} \underline{F}'(k)\underline{\lambda}(k)$$

$$+ \underline{H}'(k)\left[\underline{H}(k)\underline{P}(k)\underline{H}'(k) + \underline{R}(k)\right]^{\#}\left[\underline{z}(k)-\underline{H}(k)\hat{\underline{x}}(k|k-1)\right] \tag{D.47}$$

<u>Discussion</u>: Equations (D.40), (D.41) and (D.42) are simply other forms of (D.5), (D.6) and (D.3) respectively. The generalized inverse may be replaced by any pseudo-inverse if (D.6) is used in place of (D.41).

The calculation procedure is to use (D.40), (D.41) and (D.42) to obtain $\hat{\underline{x}}(k+1|k)$ and $\hat{\underline{x}}(k|k)$ for k=o,...,n. Then calculate $\underline{\lambda}(k)$ by backward recursion of (D.44). Once $\underline{\lambda}(k)$ is obtained, (D.43) may be used to obtain $\hat{\underline{x}}(k|n)$ for all $k \in (o,\dots,n)$.

If the measurement noise is not independent, i.e., $E[\underline{v}(j)\underline{v}'(k)] \neq \delta_{jk}\underline{R}(k)$ we may consider this noise to be the output of a linear system excited by white noise and adjoin the states of this fictitious linear system to the basic process. There is no need to assume that $\underline{F}(k)$ is non-singular.

We can express the solution of the linear smoothing problem in a simpler form by writing the expressions in terms of $\hat{\underline{x}}(k|k-1)$. Com-

bining (D.42) and (D.43) yields

$$\underline{\hat{x}}(k|n) = \underline{\hat{x}}(k|k-1) + \underline{P}(k)\underline{H}'(k)\underline{B}(k) \left[\underline{z}(k)-\underline{H}(k)\underline{\hat{x}}(k|k-1)\right]$$

$$+ \underline{P}(k) \left[\underline{F}'(k)-\underline{H}'(k)\underline{M}(k)\right] \underline{\lambda}(k)$$

This expression may be combined with (D.44) to obtain the following equation, which is a restatement of (D.35);

$$\underline{\hat{x}}(k|n) = \underline{\hat{x}}(k|k-1) + \underline{P}(k)\underline{\lambda}(k-1) \tag{D.48}$$

We may summarize this result in the following corollary;

<u>Corollary</u>: The solution of the linear smoothing problem is given by (D.48), (D.40), (D.41) and (D.44). (D.44) is now valid for k=o,...,n. The boundary condition $\underline{\lambda}(n) = \underline{o}$ remains in effect. The <u>a priori</u> mean $\underline{m}$ is to be used for $\underline{\hat{x}}(o|-1)$ in these expressions.

APPENDIX E

PROBABILITY DENSITY FUNCTIONALS FOR GAUSSIAN RANDOM PROCESSES

In this appendix we shall given an introductory discussion of probability density functionals. Our treatment will be formal, since a rigorous discussion of "white noise" and probability density functionals is beyond the scope of this report. The concept of a probability density functional was introduced by Preston [53] and was extended by Thomas and Zadeh [62] to include vector Gaussian processes. For a more abstract treatment the reader is referred to the work of Parzen [72].

Consider a vector random process $\{\underline{x}(t)\}$, for $o \leq t \leq T$, consisting of a sequence of steps occurring at intervals T/N apart. Assume that the magnitude of the steps is normally distributed with zero mean and the following correlation (covariance) matrix;

$$E\left[\underline{x}(iT/N)\underline{x}'(jT/N)\right] = \underline{R}(iT/N,jT/N) \qquad i,j \in (1,\ldots N) \qquad (E.1)$$

A sample function of a scalar version of such a process is shown in Fig. E-1.

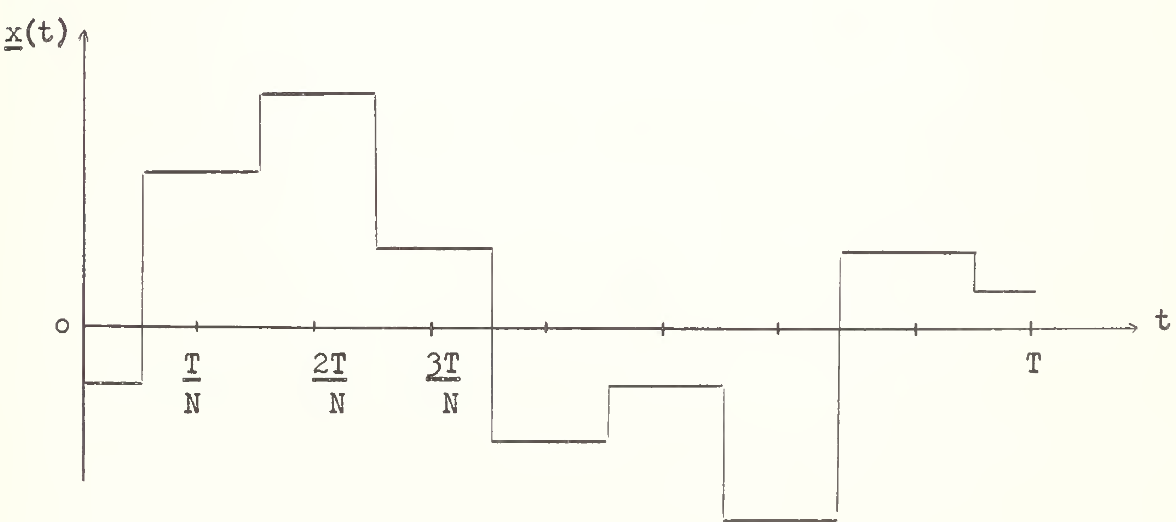


Fig. E-1

The probability density function for the sequence of points $\{\underline{x}(T/N), \underline{x}(2T/N), \ldots, \underline{x}(T)\}$ is

$$p[\underline{x}(T/N), \ldots, \underline{x}(T)] = C^{-1} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \underline{x}'(iT/N)\underline{H}(iT/N, jT/N)\underline{x}(jT/N)\right\} \tag{E.2}$$

where C is a constant of proportionality and where $\underline{H}$ and $\underline{R}$ are related by the following expression;

$$\sum_{k=1}^{N} \underline{R}(iT/N, kT/N)\underline{H}(kT/N, jT/N) = \underline{I}\ \delta_{ij} \tag{E.3}$$

Let us define a function, $\underline{G}(iT/N, jT/n)$, as follows;

$$\underline{G}(iT/N, jT/N) = \underline{H}(iT/N, jT/N)\ [N/T]^2 \tag{E.4}$$

Then (E.2) may be rewritten in the following form;

$$p[\underline{x}(T/N), \ldots, \underline{x}(T)] = C^{-1} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \underline{x}'(iT/N)\underline{G}(iT/N, jT/N)\underline{x}(jT/N)\ [T/N]^2\right\} \tag{E.5}$$

We may also rewrite (E.3) as

$$\sum_{k=1}^{N} \underline{R}(iT/N, kT/N)\underline{G}(kT/N, jT/N)\ [T/N] = \underline{I}\ \delta_{ij}\ [N/T] \tag{E.6}$$

If we take the limit as $N \rightarrow \infty$ the resulting process $\{\underline{x}(t)\}$ is a Gaussian random process. A random process $\{\underline{x}(t)\}$ is said to be Gaussian if the joint distribution for every finite set of points $\{\underline{x}(t_1), \ldots, \underline{x}(t_n)\}$ is normal.

In the limit as $N \rightarrow \infty$, (E.6) becomes

$$\int_0^T \underline{R}(t_1,t)\underline{G}(t,t_2)dt = \underline{I}\ \delta(t_1-t_2), \quad (0 \le t_1,t_2 \le T) \tag{E.7}$$

A matrix kernel $\underline{G}(t,t_2)$ satisfying (E.7) is said to be inverse to the correlation matrix $\underline{R}(t_1,t)$.

If we were to take the limit of (E.5) as $N \to \infty$, the constant of proportionality C would become infinite. To avoid this difficulty we define the probability density functional as a ratio, using a definition similar to that given by Thomas and Zadeh [62]. Let $\underline{z}_{[0,T]}$ be a vector of curves. Let h be a small positive scalar and let $\underline{h}$ be a vector, each element of which is equal to h. Then, the probability density functional of the process $\{\underline{x}(t)\}$ on the interval $[0,T]$ is defined as follows;

$$p_{\underline{x}}\left[\underline{z}_{[0,T]}\right] = \lim_{h \to 0} \frac{Pr\left[\underline{z}(t)-\underline{h} < \underline{x}(t) \le \underline{z}(t)+\underline{h},\ \text{for } 0 \le t \le T\right]}{Pr\left[-\underline{h} < \underline{x}(t) \le \underline{h},\ \text{for } 0 \le t \le T\right]} \tag{E.8}$$

$$= \exp\left\{-\frac{1}{2}\int_0^T\int_0^T \underline{z}'(t_1)\underline{G}(t_1,t_2)\underline{z}(t_2)\,dt_1\,dt_2\right\} \tag{E.9}$$

Note the similarity between (E.5) and (E.9).

White Gaussain noise is defined to be the formal limit of a Gaussian process as the values of $\{\underline{x}(t)\}$ at different instants of time become statistically independent. In this case $\underline{R}(t_1,t_2)$ becomes $\underline{R}(t_1)\ \delta(t_1-t_2)$ and $\underline{G}(t_1,t_2)$ becomes $\underline{R}^{-1}(t_1)\ \delta(t_1-t_2)$. Then (E.9) becomes

$$p_{\underline{x}}\left[\underline{z}_{[0,T]}\right] = \exp\left\{-\frac{1}{2}\int_0^T \underline{z}'(t)\underline{R}^{-1}(t)\underline{z}(t)\,dt\right\} \tag{E.10}$$

$$= \exp\left\{-\frac{1}{2}\int_0^T \|\underline{z}(t)\|^2_{\underline{R}^{-1}(t)}\,dt\right\} \tag{E.11}$$

Probability density functionals are a powerful tool when used for problems involving Gaussian random processes on a finite time interval and will be used in Chapter III of this report.

REFERENCES

1. R.C. Amara, "Application of Matrix Methods to the Linear Least Squares Synthesis of Multi-Variable Systems", J. Franklin Inst., Vol. 268, 1959, pp. 1-16.

2. A.R. Amir-Moéz and A.L. Fass, Elements of Linear Spaces, Pergamon Press, New York, 1962.

3. R.E. Bellman, Adaptive Control Processes: A Guided Tour, Princeton University Press, Princeton, New Jersey, 1961.

4. R.E. Bellman, Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1957.

5. R.E. Bellman and S.E. Dreyfus, Applied Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1962.

6. A. Ben-Isreal and S.J. Wersan, "A Least-Square Method for Computing the Generalized-Inverse of an Arbitrary Complex Matrix", O.N.R. Research Memorandum, No. 61, The Technological Institute, Northwestern University, June 1962.

7. J.E. Bertram, "The Concept of State in the Analysis of Discrete-Time Control Systems", presented at the Joint Automatic Control Conference, New York, New York, 1962.

8. H.W. Bode and C.E. Shannon, "A Simplified Derivation of Linear Least-Squares Smoothing and Prediction Theory", Proc. IRE, Vol. 38, 1950, pp. 417-425.

9. A.E. Bryson and W. Denham, "A Steepest Ascent Method for Solving Optimum Programming Problems", J. App. Mech., June 1962, pp. 247-257.

10. A.E. Bryson and M. Frazier, "Smoothing for Linear and Nonlinear Dynamic Systems", presented at the Optimum System Synthesis Conference, Wright-Patterson Air Force Base, Ohio, September 1962.

11. S.S.L. Chang, Synthesis of Optimum Control Systems, McGraw-Hill Book Company, Inc., New York, 1961.

12. W.B. Davenport, Jr. and W.L. Root, Introduction to Random Signals and Noise, McGraw-Hill Book Company, Inc., New York, 1958, pp. 239-250, 371-382.

13. M.C. Davis, "Optimum Systems in Multidimensional Random Processes", Sc.D. Thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, June 1961.

14. C.A. Desoer, "The Bang Bang Control Problem Treated by Variational Techniques", Information and Control, Vol. 2, 1959, pp. 333-348.

15. S.E. Dreyfus, "Dynamic Programming and the Calculus of Variations", Journal of Mathematical Analysis and Applications, Vol. 1, 1960, pp. 228-239.

16. S.E. Dreyfus, "Variational Problems with State Variable Inequality Constraints", RAND publication P-2605, July 1962.

17. R.A. Frazier, W.J. Duncan and A.R. Collar, Elementary Matrices and Some Applications to Dynamics and Differential Equations, Cambridge University Press, Cambridge, 1960.

18. B. Friedland, "The Structure of Optimum Control Systems", J. Basic Engr., (ASME Trans.), 84-D, 1962, 1-12.

19. A.T. Fuller, "Phase Space in the Theory of Optimum Control", J. of Electronics and Control, Vol. 8, May 1960.

20. J.E. Gibson, (Ed.), Proceedings of Dynamic Programming Workshop, Purdue University, Boulder, Colorado, June 1961.

21. T.N.E. Greville, "The Pseudo-Inverse of a Rectangular or Singular Matrix and its Applications to the Solution of a System of Linear Equations", SIAM Review 1, 1959, pp. 38-43.

22. T.N.E. Greville, "Some Applications of the Pseudo-Inverse of a Matrix", SIAM Review 2, No. 1, 1960, pp. 15-22.

23. T.L. Gunkel II, "Optimum Design of Sampled-Data Systems with Random Parameters", Stanford Electronic Laboratories Technical Report, No. 2101-2, April 1961.

24. H. Hancock, The Theory of Maxima and Minima, Dover Publications, Inc., New York, 1960.

25. Y.C. Ho, "The Method of Least Squares and Optimal Filtering Theory", RAND Memorandum, RM-3329-PR, October 1962.

26. R.A. Howard, Dynamic Programming and Markov Processes, Technology Press and John Wiley and Sons, Inc., Cambridge and New York, 1960.

27. H.C. Hsieh and C.T. Leondes, "On the Optimum Synthesis of Multiple Control Systems in the Wiener Sense", IRE PGAC Trans., Vol. Ac-4, No. 2, November 1959, pp. 16-29.

28. P.D. Joseph and J.T. Tou, "On Linear Control Theory", AIEE Paper 61-728.

29. R.E. Kalman, "New Methods and Results in Linear Prediction and Filtering Theory", RIAS TR 61-1, Baltimore, Maryland, 1961.

30. R.E. Kalman, "The Theory of Optimal Control and the Calculus of Variations", RIAS TR 61-3, Baltimore, Maryland, 1961.

31. R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", J. Basic Engr., (ASME Trans.) 82-D, 1960, 35-45.

32. R.E. Kalman, "On the General Theory of Control Systems" Proc. First IFAC Congress, Moscow, USSR, Butterworth and Co., London, England, 1961, pp. 481-496.

33. R.E. Kalman, "Mathematical Description of Linear Dynamical Systems", presented at the Symposium on Multivariable System Theory, SIAM, November 1962; to appear in SIAM series on Control, No. 2, 1963.

34. R.E. Kalman and R.S. Bucy, "New Results in Linear Filtering and Prediction Theory", J. Basic Engr., (ASME Trans.) 83-D, 1961, 95-108.

35. H.J. Kelley, "Methods and Gradients", chapter 6, Optimization Techniques, G. Leitmann (Ed.), Academic Press, New York and London, 1962.

36. W. Kipiniak, Dynamic Optimization and Control: A Variational Approach, Technology Press and John Wiley and Sons, Inc., Cambridge and New York, 1961.

37. R.E. Kopp and R.J. Orford, "Linear Regression Applied to System Identification for Adaptive Control Systems", presented at American Rocket Society 17th Annual Meeting and Space Flight Exposition, Los Angeles, California, November 1962.

38. H.J. Kushner, "Hill Climbing Methods for the Optimization of Multi-Parameter Noise Disturbed Systems", presented at the Joint Automatic Control Conference, New York, New York, 1962.

39. J.H. Laning, Jr. and R.H. Battin, Random Processes in Automatic Control, McGraw-Hill Book Company, Inc., New York, 1956.

40. J.P. LaSalle and S. Lefschetz, Stability by Liapunov's Direct Method, Academic Press, New York and London, 1961.

41. J.P. LaSalle and S. Lefschetz, (Eds.), International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics, Academic Press, New York and London, 1963.

42. Y.W. Lee, Statistical Theory of Communication, John Wiley and Sons, Inc., New York, 1960.

43. M. Margolis and C.T. Leondes, "On the Theory of Adaptive Control Systems; the Learning Model Approach", Proc. First IFAC Congress, Moscow, USSR, Butterworth and Co., London, England. 1961, pp. 555-563.

44. C.W. Merriam III, "An Optimization Theory for Feedback Control System Design", Information and Control, Vol. 3, 1960, pp. 32-59.

45. C.W. Merriam III, "Mathematical Error Criteria in the Design of Adaptive Control Systems", Trans. AIEE, Applications and Industry, January 1960.

46. E. Mishkin and L. Braun, (Eds.), Adaptive Control Systems, McGraw-Hill Book Company, New York, 1961.

47. E.H. Moore, Bull. Amer. Math. Soc. 26, 1920, pp. 394-395.

48. E.H. Moore, General Analysis, Part I, Mem. Amer. Philos. Soc. I, 1935.

49. G.C. Newton, Jr., L.A. Gould and J.F. Kaiser, Analytical Design of Linear Feedback Controls, John Wiley and Sons, Inc., New York, 1957.

50. R. Penrose, "A Generalized Inverse for Matrices", Proc. Camb. Philos. Soc. 51, 3, 1955, pp. 406-413.

51. R. Penrose, "On Best Approximate Solution of Linear Matrix Equations", Proc. Camb. Philos. Soc. 52, I, 1956, pp. 17-19.

52. L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze and E.F. Mischenko, The Mathematical Theory of Optimal Processes, Interscience, New York, 1962.

53. G.W. Preston, "Equivalence of Optimum Transducers and Sufficient and Most Efficient Statistics", J. of App. Physics, Vol. 24, No. 7, July 1953, pp. 841-844.

54. H. Raiffa and R. Schlaifer, Applied Statistical Decision Theory, Division of Research Graduate School of Business Administration, Harvard University, Boston, 1961.

55. L.I. Rozonoer, "L.S. Pontryagin's Maximum Principle in the Theory of Optimum Systems", Automation and Remote Control, Vol. 20, Oct., Nov., Dec., 1959, pp. 1288-1302, pp. 1405-1421, pp 1517-1532.

56. D.J. Sakrison, "Application of Stochastic Approximation Methods to System Optimization", Technical Report 391, Research Laboratory of Electronics, M.I.T., Cambridge, Massachusetts, July 1962.

57. W.W. Seifert and C.W. Steeg, Control Systems Engineering, McGraw-Hill Book Company, Inc., New York, 1960.

58. S. Sherman, "A Theorem on Convex Sets with Applications", Ann. Math. Stat., 26, 1955, pp. 763-767.

59. G.L. Smith, "Multivariable Linear Filter Theory Applied to Space Vehicle Guidance", presented at SIAM Symposium on Multivariable System Theory, Cambridge, Massachusetts, November 1962.

60. V.V. Solodovnikov, Introduction to the Statistical Dynamics of Automatic Control, Dover Publications, Inc., New York, 1960.

61. D. ter Haar, Elements of Hamiltonian Mechanics, North-Holland Publishing Company, Amsterdam, 1961.

62. J.B. Thomas and L.A. Zadeh, "Note on an Integral Equation Occurring in the Prediction, Detection and Analysis of Multiple Time Series", IRE Trans. on Information Theory, Vol. IT-7, No. 2, April 1961, pp. 118-120.

63. J.G. Truxall, Automatic Feedback Control System Synthesis, McGraw-Hill Book Company, Inc., New York, 1955.

64. J.H. Westcott, (Ed.), An Exposition of Adaptive Control, Pergamon Press, New York, 1962.

65. N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications, The University Press of the Massachusetts Institute of Technology, Cambridge, Massachusetts, 1949.

66. N. Wiener, Nonlinear Problems in Random Theory, The University Press of the Massachusetts Institute of Technology, Cambridge, Massachusetts, 1958.

67. L.A. Zadeh and J.R. Ragazzini, "An Extension of Wiener's Theory of Prediction", J. App. Phys., Vol. 21, 1950, pp. 645-655.

68. L.A. Zadeh, "An Introduction to State-Space Techniques", presented at the Joint Automatic Control Conference, New York, New York, 1962.

69. J.E. Bertram and P.E. Sarachik, "On Optimum Computer Control", Proc. First IFAC Congress, Moscow, USSR, Butterworth and Co., London, England, 1961, pp. 419-422.

70. R.E. Kalman, Y.C. Ho and K.S. Narendra, "Controllability of Linear Dynamical Systems", Contributions to Differential Equations, Vol. I, (to appear).

71. L. Markus and E.B. Lee, "On the Existence of Optimal Controls", J. Basic Engr., (ASME Trans.) Vol. 84-D, 1962, pp. 13-22.

72. E. Parzen, "Extraction and Detection Problems and Reproducing Kernel Hilbert Spaces", J. SIAM Control, Series A, Vol. I, No.1, 1962, pp. 35-62.

BIOGRAPHICAL NOTE

Henry Cox was born in Philadelphia, Pennsylvania on March 7, 1935. He is married to the former Mary Ann Shaw of Brooklyn, New York.

Lieutenant Cox is a graduate of Brooklyn Preparatory School. He attended the College of the Holy Cross, Worcester, Massachusetts, where he participated in the regular NROTC program, graduating in 1956 with the Bachelor of Science degree in physics.

Upon graduation, he served for one year aboard USS Pocono (AGC-16). He was the Engineer Officer of USS Kleinsmith (APD-134) from 1957 until 1959. He then reported to the Massachusetts Institute of Technology.

Lieutenant Cox will be designated as an Engineering Duty officer upon graduation. He is a member of the Institute of Electrical and Electronics Engineers, the Society of Naval Architects and Marine Engineers, Tau Beta Pi and Sigma Xi.